DTIC FILE COPY

AFOSR-TR- 88-1001

(2)

AD-A200 892

RF Project 763265/714837
Final Report

EFFECTS OF ASSUMING INDEPENDENT COMPONENT FAILURE TIMES, IF THEY ARE ACTUALLY DEPENDENT, IN A SERIES SYSTEM

Melvin L. Moeschberger
Department of Preventive Medicine
and
John P. Klein
Department of Statistics

For the Period
September 1, 1982 - December 31, 1987

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH/PKZ
Bolling AFB, D.C. 20332

Contract No. AFOSR-82-0307

May 1988

OSU

The Ohio State University
Research Foundation
1314 Kinnear Road
Columbus, Ohio 43212

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY: NA	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for Public Release; Distribution Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE: NA	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 88-1001
6a. NAME OF PERFORMING ORGANIZATION: The Ohio State University; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION: AFOSR/NM
6c. ADDRESS (City, State and ZIP Code): 1314 Kinnear Road, Columbus, Ohio 43212	7b. ADDRESS (City, State and ZIP Code): Bldg. 410, Bolling AFB, DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR; 8b. OFFICE SYMBOL (If applicable): NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: AFOSR-82-0307
8c. ADDRESS (City, State and ZIP Code): Bldg. 410, Bolling AFB, DC	10. SOURCE OF FUNDING NOS. — PROGRAM ELEMENT NO.: 6.1102F; PROJECT NO.: 2304; TASK NO.: ; WORK UNIT NO.:

11. TITLE (Include Security Classification) Effects of Assuming Independent Component Failure Times, if They are Actually Dependent in a Series System

12. PERSONAL AUTHOR(S) Melvin L. Moeschberger and John P. Klein

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Final	FROM 9-1-82 TO 12-31-87	May 31, 1988	116

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	modeling series systems, system reliability, competing risks bivariate exponential distributions, test for independence, consistency of the product limit estimator (Continued)
XXXXXXXXXXXXXXX			

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The overall objective of this proposal is to develop improved estimation techniques for use in reliability studies when there are competing failure modes or competing causes of failure associated with a single failure mode in data from series systems. Such improved nonparametric estimators of the component failure distributions will be accomplished by incorporating some dependence structure between the potential component failure times. The first specific aim is to investigate techniques which identify departures from independence, based on data collected from series systems, by making some restrictive assumption about the structure of the system, and obtain modified nonparametric estimators which incorporate some restrictive assumptions about the structure of the system. The second aim will be to develop improved nonparametric estimators of component reliability based on data from a series system with independent component lifetimes by obtaining modifications of the product limit estimator which incorporate some parametric information and by studying the robustness (Continued)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL: Brian W. Woodruff, Major	22b. TELEPHONE NUMBER (Include Area Code): (202) 767-5027; 22c. OFFICE SYMBOL: AFOSR/NM

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

Block 18- Subject terms (Continued)

dependent risks, estimation of component life

Block 19- Abstract (Continued)

of these estimators to misspecification of the parametric model. Competing risk analyses have been performed in the past and will continue to be performed in the future. This study will provide the user of such techniques with an alternative to the usual approach of assuming independent risks, an assumption which most of the methods currently in use assume.

EFFECTS OF ASSUMING INDEPENDENT COMPONENT FAILURE TIMES, IF THEY ARE ACTUALLY DEPENDENT, IN A SERIES SYSTEM

Melvin L. Moeschberger
Department of Preventive Medicine
and
John P. Klein
Department of Statistics

For the Period
September 1, 1982 - December 31, 1987

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH/PKZ
Bolling AFB, D.C. 20332

Contract No. AFOSR-82-0307

May 1988

Table of Contents

A. TECHNICAL SECTION

1. Abstract

The overall objective of this proposal is to develop improved estimation techniques for use in reliability studies when there are competing failure modes or competing causes of failure associated with a single failure mode in data from series systems. Such improved nonparametric estimators of the component failure distribution will be accomplished by incorporating some dependence structure between the potential component failure times. The first specific aim is to investigate techniques which identify departures from independence, based on data collected from series systems, by making some restrictive assumption about the structure of the system, and obtain modified nonparametric estimators which incorporate some restrictive assumptions about the structure of the system. The second aim will be to develop improved nonparametric estimators of component lifetimes by obtaining modifications of the product limit estimator which incorporate some parametric information and by studying the robustness of these estimators to misspecification of the parametric model. Competing risk analyses have been performed in the past and will continue to be performed in the future. This study will provide the user of such techniques with an alternative to the usual approach of assuming independent risks, an assumption which most of the methods currently in use assume.

2. Objectives

The overall objective of this proposal is to continue our investigation into improved estimation techniques for use in reliability studies when there are competing failure modes or competing causes of failure associated with a single failure mode in data from series systems. We shall term such experiments as competing risks experiments. Our primary goal is to obtain improved nonparametric estimators of the component failure distributions incorporating some dependence structure between the potential component failure times. The specific objectives are to continue our investigation into:

(1) the problems associated with dependent systems by

- (a) investigating techniques which identify departures from independence, based on data collected from series systems, by making some restrictive assumption about the structure of the system, and
- (b) obtaining modified nonparametric estimators which incorporate some restrictive assumptions about the structure of the systems,

(2) improved nonparametric estimators of component reliability based on data from a series system with independent component lifetimes by

- (a) obtaining modifications of the product limit estimator which incorporate some parametric information, and
- (b) studying the robustness of these estimators to misspecification of the parametric model.

3. Introduction to Problem and Significance of Study

Alvin Weinberg (1978) in an editorial comment in the published proceedings of a workshop on Environmental Biological Hazards and Competing risks noted that "the question of competing risks will not quietly go away: corrections for competing risks should be applied routinely to data." The problem of competing risks commonly arises in a wide range of experimental situations. Although we shall confine our attention in the following discussion to those situations involving series systems in which competing failure modes or competing causes of failure associated with a single mode are present, it is certainly true that we might just as easily speak of clinical trials, animal experiments, or other medical and biological studies where competing events interrupt our study of the main event of interest (cf. Lagakos (1979)).

Consider electronic or mechanical systems, such as satellite transmission equipment, computers, aircraft, missiles and other weaponry consisting of several components in series. Usually each component will have a random life length and the life of the entire system will end with the failure of the shortest lived component.

Competing risks arise in such reliability studies when

1) the study is terminated due to a lack of funds or the pre-determined period of observation has expired (Type I censoring).

2) the study is terminated due to a pre-determined number of failures of the particular failure mode of interest being observed (Type II censoring).

3) some systems fail because components other than the one of interest malfunction.

4) the component of interest fails from some cause other than the one of interest.

In all four situations, one may think of the main event of interest as being censored, i.e., not fully observable. In the first two situations, the time to occurrence of the event of interest should be independent of the censoring mechanism. In such instances, the methodology for estimating relevant reliability probabilities has received considerable attention (cf. David and Moeschberger (1978), Kalbfeish and Prentice (1980), Elandt-Johnson and Johnson (1980), Mann, Schafer, Singpurwalla (1974) and Barlow and Proschan (1975) for references and discussion). In the third situation, the time to failure of the component of interest may or may not be independent of the failure times of other components in the system. For example, there may be common environmental factors such as extreme temperature which may affect the lifetime of several components. Thus the question of dependent competing risks is raised. A similar observation may be made with respect to the fourth situation, viz., failure times associated with different failure modes of a single component may be dependent.

In our earlier work we have demonstrated that one can be appreciably misled if one assumes independent component lifetimes when they are really dependent. One purpose of this renewal is to explore improved estimation techniques which incorporate some dependence structure between the potential component failure times. Another aim is to investigate techniques which identify departures from independence. A third aim is to obtain modifications of the product lim. estimator in the presence of independent censoring which incorporate some parametric information and to study the robustness of these estimators to misspecification of the parametric information.

In summary, competing risk analyses have been performed in the past and will continue to be performed in the future. This study will provide the user of such techniques with an alternative to the usual approach of assuming independent risks, an assumption which most of the methods currently in use assume.

4. Report of the Work During 9-1-82 to 12-31-87

We shall briefly review the work performed during September 1, 1982 to December 31, 1987 under Air Force Office of Scientific Research grant number 82-0307 by providing a summary of the scientific manuscripts arising from this work, and stating the publication status of such papers.

First we believe that substantial progress has been made in assessing the error in modeling system life in a series system assumed to have independent component lifetimes when, in fact, the component lifetimes are dependent. The results of our study of the effect of erroneously assuming independence is summarized in four papers. We have investigated in detail the effects of this assumption for several multivariate exponential distributions. Results for multivariate Weibull and gamma distributions will be similar since these can be obtained from the exponential by simple marginal transformations.

The first paper deals with the model proposed by Gumbel (1960).

Moeschberger, M.L. and Klein, J.P. (1984). Consequences of departures from independence in exponential series systems. Technometrics, 26: 277-284. (Appendix A). This paper considers the model

$$P(X > x, Y > y) = \exp(-\lambda_1 x - \lambda_2 y)\,[1 + \alpha(1-\exp(-\lambda_1))(1-\exp(-\lambda_2 y))].$$

The effects of erroneously assuming independence on modeling system reliability and system mean life are examined as well as the effects of erroneously assuming independence on the Mann-Grubbs (1974) confidence bounds on system reliability.

A second paper investigates the effects of erroneously assuming independence in both parametric estimation of component parameters and in the Kaplan-Meier (1958) product limit estimator of the component reliability.

Klein, J.P. and Moeschberger, M.L. (1984). Asymptotic bias of the product limit estimator under dependent competing risks. Indian Journal of Productivity, Reliability, and Quality Control, 9, 1-7. (Appendix B) In this paper it is shown that for a general dependence structure the product limit estimator is not consistent but converges to another marginal distribution which can be expressed in terms of the system reliability and the crude system reliability. When the risks are dependent and fall in the constant sum class of Williams and Lagakos (1977) then the estimator is consistent.

A third paper investigates the effect of the independence assumption in both series and parallel systems for the Marshall-Olkin (1967) distribution.

Klein, J.P. and Moeschberger, M.L. (1986). The independence assumption for a series or parallel system when component lifetimes are exponential. IEEE Transactions on Reliability, Vol. R-35, No. 3, 330-335. (Appendix C) This paper shows that for the Marshall-Olkin (1967) model the error in predicting mean system life can be as large as 100% of the mean system life under independence and the error in modeling system reliability, which depends upon the mission time, can be as large as 200% for large mission times.

A fourth paper compares the effects of the independence assumption for the three models of Gumbel (1960), the model of Downton (1970), and the model of Oakes (1982).

Klein, J.P. and Moeschberger, M.L. (1987). Independent or dependent competing risks? Does it make a difference? Communications in Statistics Issue B-16, No. 2, 507-533. (Appendix D) This paper considers errors in modeling system mean life and system reliability for the above models. It also examines the models with maximal and minimal correlations (Frechet (1951)) and obtains bounds on the possible error one can incur in modeling system mean life or system reliability.

These papers show that even for relatively small correlation there is an appreciable estimation error one incurs in estimating the parameters of the components.

Another line of research on the grant was to derive bounds on component reliability when the failure models are dependent and fall in a particular dependence class. The details for our approach are found in the following paper

Klein, J.P. and Moeschberger, M.L. (1988). Bounds on net survival probabilities for dependent competing risks. (To appear in Biometrics) (Appendix E). This paper obtains bounds on the component reliability, based on data from a series system, for the Oakes (1982) model. Since this model has the same dependence structure as a random effects model with w have a gamma distribution, these bounds are good for a general class of distributions. The bounds, which are determined by specification of a range of coefficients of concordance, are found by solving a differential equation in the observable system reliability and crude life on one hand and the unobservable component survival function on the other hand.

A consequence of the previous research was the development of a test for independence when the component reliabilities are known. Details of this work are found in the following paper

Klein, J.P. (1986). A test for independence based on data from a series system, Reliability and Quality Control, A.P. Basu, ed., 235-244. North-Holland. (Appendix F). This paper provides a modification of Kendall's test, based on the coefficient of concordance, for data from a series system. The test uses the component survival probabilities to partially estimate the probability of concordance or discordance. A Monte Carlo study shows that this test has reasonable power for several underlying models of dependence.

An additional line of research has consisted of examining the problem of improving the product-limit estimator of Kaplan and Meier (1958) when there is extreme independent right censoring. The results are summarized in

Moeschberger, M.L. and Klein, J.P. (1985). A comparison of several methods of estimating the survival function when there is extreme right censoring. Biometrics, 41, 253-260 (Appendix G). This paper looks at several techniques for completing the product-limit estimator by estimating the tail probability of the survival curve beyond the largest observed death time. Two methods are found to work well for a variety of underlying distributions. The first method replaces those censored observations larger than the biggest death time by the expected order statistics, conditional on the largest death, computed from a Weibull distribution. The Weibull is chosen since it is known to be a reasonable model for survival in many situations. Parameters of the model are estimated in several ways, but the method of maximum likelihood seems to provide the best results. The second method replaces the constant value of the product-limit estimator beyond the last death time by the tail of a Weibull survival function. Again parameters are estimated by a variety of methods with the maximum likelihood estimators performing the best.

A second paper which has been developed along similar lines as the one preceding follows.

Klein, J.P., Lee, Shin Chang, and Moeschberger, M.L. (1987). A partially parametric estimator of survival in the presence of randomly censored data. (Currently being revised for publication) (Appendix H). This paper suggest an improvement of the Kaplan-Meier product-limit estimator when the censoring mechanism is random. The proposed estimator treats the uncensored observations nonparametrically and uses a parametric model only for the censored observations. One version of this proposed estimator always has a small bias and mean-squared error than the product-limit estimator. An example estimating the survival function of patients enrolled in The Ohio State University Bone Marrow Transplant Program is presented.

Another line of research has been developed which discusses some general properties of a random environmental stress model. Suppose that under ideal conditions such as one might find in the laboratory testing stage of development, the component hazard rates are $h_1(t),\ldots,h_p(t)$ and that the component lifetimes of the p components in the series system are independent. When the system is put into use under field conditions, there is a common environmental factor which simultaneously changes each component's hazard rate to $wh_1(t)$, $wh_2(t),\ldots,wh_p(t)$. We have investigated this model when component lifetimes under independence are exponential and w has a variety of distributions including the uniform and gamma distribution (Sukhoon Lee's Ph.D. thesis under Dr. Klein). Also we have studied this model when the components, under ideal conditions, are of a Weibull form and w has a gamma distribution. Estimation of parameters under the gamma stress model is considered, and a new estimator based on scaled total time on test transform is presented. These results were reported in a series of papers.

Klein, John P. and Lee, S. (1985). "On dependent competing risks." Contributed Papers, 45th Session of the International Statistical Institute, Book 1, 263-264. This paper surveyed the random environmental stress model for series and parallel systems focusing on the robustness of independence assumption in modeling series and parallel systems.

Lee, Sukhoon and Klein, John P. (1987). "Bivariate models with a random environmental factor." IAPQR Transactions (To appear) (Appendix I). Studies the probabilistic properties of the random environmental stress model. General results characterizing the dependence structure are obtained and several specific examples are considered.

Lee, Sukhoon, and Klein, John P. (1987) "Statistical methods for combining laboratory and field data based on a random environmental stress model." (Submitted for publication) (Appendix J). In this paper we assume the environmental stress model with exponential distributions for the components under ideal conditions and w having a gamma distribution. The type of data available consists of component data collected under ideal conditions and system failure data collected under operating conditions. Maximum likelihood and method of moments estimation of model parameters is considered as well as a least squares estimator based on the total time on test transform. The problem of experimental design is also considered in detail.

A final paper in this series is

Klein, J.P. and Lee, Sukhoon. (1986). A random environmental stress model for competing risks. (Submitted for publication) (Appendix K). This paper, which was presented at the 1986 Missouri Conference on Reliability and Quality Control surveys the results reported in the above aspects.

Another paper which was co-sponsored by this grant is

Klein, J.P. and Basu, A.P. (1985). Estimating reliability for bivariate exponential distributions, Sankhya B:47, 346-353. (Appendix L). This paper considers the problem of estimating reliability for the bivariate distributions of Block and Basu (1974) and Marshall and Olkin (1967). For the Block-Basu model, a minimum variance unbiased estimator of the joint survival function is obtained in the case of identically distributed marginals. For the non-identically distributed case, the performance of the maximum likelihood estimator and the jackknifed maximum likelihood estimator is studied. For the Marshall-Olkin model, the performance of several different parameter estimators and bias reduction techniques for estimation of joint reliability are considered.

Finally, the most recent manuscript prepared under this grant was

Klein, J.P. and Moeschberger, M.L. (1988). The robustness of several estimators of the survivorship function with randomly censored data. (Submitted for publication). (Appendix M). This paper studies the efficiency of the Kaplan-Meier and the fully parametric approach in estimating the survivorship function when a particular model such as the exponential, Weibull, normal, log normal, exponential power, Pareto, Gompertz, gamma, or bathtub shaped hazard distributions is assumed under a variety of censoring schemes and underlying failure models. We conclude that in most cases the parametric estimators outperform the distribution free estimator. The results are particularly striking if the Weibull form of these estimators are used routinely.

All the results found in the preceding papers have been presented to regional, national, and international statistics, reliability, and quality control meetings.

References

Barlow, R.E. and Proschan, F. (1975). Statistical Theory of Reliability and Life Testing - Probability Models. Holt, Rinehart and Winston, Inc., New York.

Block, H.W. and Basu, A.P. (1974). A continuous bivariate exponential extension. J. Amer. Statist. Assoc. 69, 1031-37.

David, H.A. and Moeschberger, M.L. (1978). Theory of Competing Risks. Griffin, London.

Downton, F. (1970). Bivariate exponential distributions in reliability theory. J. Roy. Statist. Soc. B. 32: 408-417.

Elandt-Johnson, R.C. and Johnson, N.L. (1980). Survival Models and Data Analysis. Wiley, New York.

Frechet, M. (1951). Sur les Tableaux de Correlatoin Dont les Marges Sont Donnees. Annales de l'Universite de Lyon. Section A, Series 3, 14, 53-77.

Gumbel, E.J. (1960). Bivariate exponential distributions. J. Amer. Statist. Assoc. 55: 698-707.

Kalbfleisch, J.D. and Prentice, R.L. (1980). The Statistical Analysis of Failure Time Data. Wiley, New York.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457-481.

Lagakos, S.W. (1979). General right censoring and its impact on the analysis of survival data. Biometrics 35: 139-156.

Mann, N.R. and Grubbs, F.E. (1974). Approximately optimum confidence bounds for system reliability based on component test data. Technometrics 16: 335-347.

Marshall, A.W. and Olkin, I. (1967). A multivariate exponential distribution. J. Amer. Statist. Assoc. 66, 30-40.

Oakes, D. (1982). A model for association in bivariate survival data. J. Roy. Statist. Soc. 44, 414-422.

Weinberg, Alvin (1978). Editorial. Environmental International 1:285-287.

Williams, J.S. and Lagakos, S.W. (1977). Models for censored survival analysis: Constant-sum and variable-sum models. Biometrika 64: 215-224.

APPENDIX A

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY NA 2b. DECLASSIFICATION / DOWNGRADING SCHEDULE NA	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION The Ohio State University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NM
6c. ADDRESS (City, State, and ZIP Code) 1314 Kinnear Road Columbus, Ohio 43212	7b. ADDRESS (City, State, and ZIP Code) AFOSR/NM Bldg 410 Bolling AFB DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-82-0307
8c. ADDRESS (City, State, and ZIP Code) AFOSR/NM Bldg 410 Bolling AFB DC 20332-6448	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO.: 61102F PROJECT NO.: 2304 TASK NO.: WORK UNIT ACCESSION NO.:

11. TITLE (Include Security Classification) Consequences of Departures From Independence in Exponential Series Systems

12. PERSONAL AUTHOR(S)
M.L. Moeschberger and John P. Klein

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 9-1-82 TO 12-31-87	May 31, 1988	8

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	competing risks; component life; modeling series systems; robustness studies; system reliability; Gumbel bivariate exponential
XXXX	XXXXXXXXXX	XXX	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This article investigates the consequences of departures from independence when the component lifetimes in a series system are exponentially distributed. Such departures are studied when the joint distribution is assumed to follow a Gumbel bivariate exponential model. Two distinct situations are considered. First, in theoretical modeling of series systems, when the distribution of the component lifetimes is assumed, one wishes to compute system reliability and mean system life. Second, errors in parametric and nonparametric estimation of component reliability and component mean life are studied based on life-test data collected on series systems when the assumption of independence is made erroneously. Systems with two components are studied.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Maj. Brian Woodruff	22b. TELEPHONE (Include Area Code) (202) 767-5027 22c. OFFICE SYMBOL NM

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

TECHNOMETRICS ©, VOL. 26, NO. 3, AUGUST 1984

Consequences of Departures From Independence in Exponential Series Systems

M. L. Moeschberger
Department of Preventive Medicine
The Ohio State University
Columbus, OH 43210

John P. Klein
Department of Statistics
The Ohio State University
Columbus, OH 43210

This article investigates the consequences of departures from independence when the component lifetimes in a series system are exponentially distributed. Such departures are studied when the joint distribution is assumed to follow a Gumbel bivariate exponential model. Two distinct situations are considered. First, in theoretical modeling of series systems, when the distribution of the component lifetimes is assumed, one wishes to compute system reliability and mean system life. Second, errors in parametric and nonparametric estimation of component reliability and component mean life are studied based on life-test data collected on series systems when the assumption of independence is made erroneously. Systems with two components are studied.

KEY WORDS: Competing risks; Component life; Modeling series systems; Robustness studies; System reliability; Gumbel bivariate exponential.

1. INTRODUCTION

Consider a system consisting of several components linked in series. For such a system the failure of any one of the components causes the system to fail. Common assumptions made in modeling and analyzing data from such a system are that the component lifetimes are independent and exponentially distributed. Many authors have considered the problem of analyzing a series system with exponential component lives. For example, confidence bounds for system reliability assuming independent exponentially distributed component lifetimes were presented in Mann (1974) and Mann and Grubbs (1974). (See Mann, Schafer, and Singpurwalla 1974 for a more comprehensive review.) More recently, work invoking the assumption of independent exponentially distributed lifetimes has been presented by Chao (1981) and Miyamura (1982). Estimation of component parameters from series system data has been treated by Boardman and Kendell (1970) in the context of independent exponential component lives. Some authors suggest a nonparametric alternative to the estimation of component reliability based on series system data (compare Kalbfleisch and Prentice 1980 and Lawless 1982).

The assumption of independence is essential to these analyses and an important concern. Several authors have shown that this assumption, by itself, is not testable because based on data from a series system, there is no way to distinguish between an independent and a dependent model. (See Tsiatis 1975, Peterson 1976, and Basu 1981 for a discussion of nonidentifiability results.) In many situations one may be appreciably misled by the independence assumption.

Lagakos (1979), in a study of the effects of various types of dependence among component lifetimes, notes that most methods of analysis have assumed noninformative models of which independence is a special case. He points out, "it is important to be aware of the possible consequences of making this assumption when it is false" (p. 152). Furthermore, Easterling (1980) states in his review of Birnbaum's (1979) monograph on competing risks, "there seems to be a need for some robustness studies. How far might one be off, quantitatively, if his analysis is based on incorrect assumptions?" (p. 131).

In this article we consider the consequences of departures from independence when the component lifetimes are exponentially distributed. Such departures may be related to some common environmental factor that is present only when the components are linked together in series. The load each component is subject to is either reduced or increased according to the age of the system. To study such departures, we have selected a model proposed by Gumbel (1960). Gumbel's model retains the assumption of exponentially distributed component lifetimes while allowing the flexibility of both positive or negative mild correlation between component lifetimes.

The effects of a departure from the assumption of independent component lifetimes in a series system will be addressed for two distinct situations. The first situation arises in modeling the performance of a theoretical series system constructed from two components whose lifetimes are exponentially distributed. Here, based on testing each component separately or on engineering design principles, it is reasonable to assume that the components are exponentially distributed with known parameter values. Based on this information, we wish to calculate parameters such as the mean life or reliability of a series system constructed from these components. In Section 2 we describe how the values of these quantities are affected by departures from independence when the component parameters are completely specified. In Section 3 we study the performance of the Mann-Grubbs (1974) confidence bounds on system reliability for small sample sizes and for varying degrees of correlation, when the component parameters are estimated from component data.

The second situation involves making inferences about component lifetime distributions, reliabilities, and mean lives from data collected on series systems. Commonly, data collected on such systems are analyzed by assuming a constant-sum model, of which independence is a special case (compare Williams and Lagakos 1977 and Lagakos and Williams 1978). In Section 4 we study the properties of the maximum likelihood estimators of component parameters calculated under an assumption of independent exponential component lifetimes when the component lifetimes are Gumbel bivariate exponential. Because of the widespread use of nonparametric estimates of component reliability, we also present in Section 5 the estimation error of the Kaplan-Meier (1958) estimator when the assumption of independence is made erroneously.

2. MODELING SYSTEM RELIABILITY FROM COMPLETE COMPONENT INFORMATION

Consider a two-component series system with component life lengths X_1, X_2. Suppose that X_i has an exponential survival function

$$\bar{F}_i(t) = P(X_i > t) = \exp(-\lambda_i t),$$
$$\lambda_i, \quad t > 0, \quad i = 1, 2.$$

This assumption is made on the basis of extensive testing of each component separately or on knowledge of the underlying mechanism of failure. The value of λ_i is assumed known. If X_1, X_2 are independent, then the time to system failure has an exponential distribution with failure rate $\lambda = \lambda_1 + \lambda_2$, and the system reliability is given by

$$\bar{F}_I(t) = P[\min(X_1, X_2) > t \mid \text{independence}]$$
$$= \exp(-\lambda t). \tag{2.1}$$

Suppose that the actual joint distribution of (X_1, X_2) has the form proposed by Gumbel (1960), namely,

$$P(X_1 > x_1, X_2 > x_2) = [\exp(-\lambda_1 x_1 - \lambda_2 x_2)]$$
$$\times [1 + \alpha(1 - \exp(-\lambda_1 x_1))(1 - \exp(-\lambda_2 x_2))]. \tag{2.2}$$

The joint probability density of (X_1, X_2) is

$$f(x_1, x_2) = \lambda_1\lambda_2[\exp(-\lambda_1 x_1 - \lambda_2 x_2)]$$
$$\times [1 + \alpha(2\exp(-\lambda_1 x_1) - 1)$$
$$\times (2\exp(-\lambda_2 x_2) - 1)]. \tag{2.3}$$

where in both (2.2) and (2.3), $x_1, x_2, \lambda_1, \lambda_2 > 0$, $-1 \le \alpha \le 1$. This distribution has marginal survival functions equivalent to those for the independent model, which, in part, is the reason for choosing it. The correlation between X_1, X_2 is $\rho = \alpha/4$, and $\alpha = 0$ is equivalent to X_1, X_2 being independent. For $\rho > 0$ (< 0) the components are positively (negatively) quadrant-dependent (see Barlow and Proschan 1975). Furthermore, the conditional expectation of X_1, given $X_2 = x_2$, is

$$E(X_1 \mid X_2 = x_2) = \frac{1}{\lambda_1}[1 + 2\rho - 4\rho\exp(-\lambda_2 x_2)].$$

If (X_1, X_2) have the joint distribution (2.3), then the true system reliability is

$$\bar{F}_D(t) = P[\min(X_1, X_2) > t \mid \text{dependence}]$$
$$= \exp(-\lambda t)[1 + 4\rho(1 - \exp(-\lambda_1 t))$$
$$\times (1 - \exp(-\lambda_2 t))]. \tag{2.4}$$

From (2.1) and (2.4) we see that the error in modeling system reliability is

$$\Delta(t) = \bar{F}_D(t) - \bar{F}_I(t)$$
$$= 4\rho[1 - \exp(-\lambda_1 t)][1 - \exp(-\lambda_2 t)]$$
$$\times \exp(-(\lambda_1 + \lambda_2)t). \tag{2.5}$$

Note that $|\Delta(t)|$ increases as $|\rho|$ increases, for fixed λ_1, λ_2, and t. The magnitude of $\Delta(t)$, of course, depends on λ_1, λ_2, t, and ρ. When $\lambda_1 = \lambda_2 = \phi$, one can show that $\Delta(t)$ is maximized at $t = [\ln 2]/\phi$ (fixing ρ and ϕ). The value of $|\Delta(t)|$ at this point is $|\rho|/4$, which is at most 1/16. Representative values of $\bar{F}_D(t)$ for $\lambda_1 = 1$, $\lambda_2 = 1.5$, and $\rho = -.25, -.125, 0, .125$, and .25 are plotted in Figure 1. The curve with $\rho = 0$ corresponds to the system reliability if the assumption of independence is true. Since most applications of interest involve reliabilities of .75 or greater, in Figure 2 we plot the ratio of the 100 pth upper percentiles under dependence and independence versus the correlation. From Figure 2 it appears that when the predicted system reliability under independence is greater than .90, misspecifying the dependence parameter has little effect. In the range where the predicted system reliability under independence is less than .75, however, misspecifying the de-

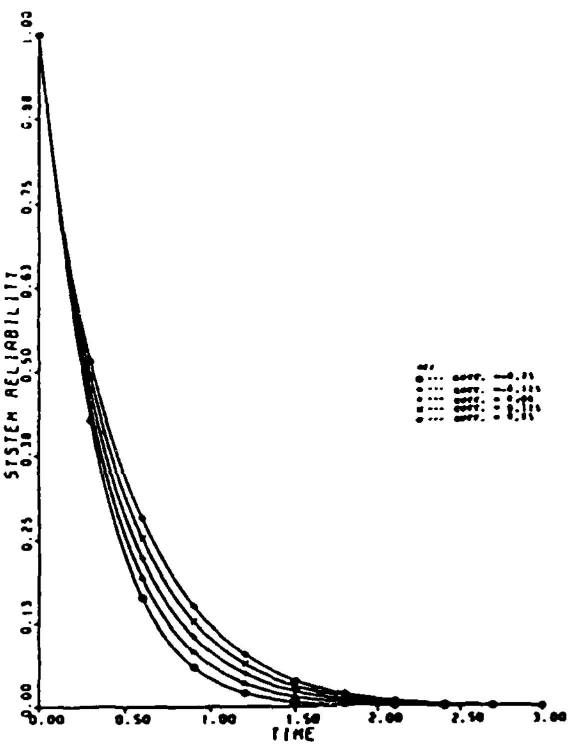


Figure 1. *System Reliability for Gumbel's Model,* $\lambda_1 = 1, \lambda_2 = 1.5$.

pendence parameter may lead to errors exceeding 6%. Maximum values of $|\Delta(t)|$ are presented in Table 1 for $\lambda_1 = 1$ and various values of λ_2.

The mean time to system failure based on (2.1), assuming independence, is

$$\mu_I = 1/(\lambda_1 + \lambda_2), \tag{2.6}$$

and that based on (2.4) is

$$\mu_D = \frac{1}{(\lambda_1 + \lambda_2)} + 4\rho\left[\frac{3}{2(\lambda_1 + \lambda_2)} - \frac{1}{(2\lambda_1 + \lambda_2)} - \frac{1}{(\lambda_1 + 2\lambda_2)}\right]. \tag{2.7}$$

The amount of error in modeling system mean life is

$$\mu_D - \mu_I = \frac{6\rho\lambda_1\lambda_2}{(\lambda_1 + \lambda_2)(2\lambda_1 + \lambda_2)(\lambda_1 + 2\lambda_2)} = \frac{6\rho\lambda_1\lambda_2\mu_I}{(2\lambda_1 + \lambda_2)(\lambda_1 + 2\lambda_2)}, \tag{2.8}$$

whose absolute value obviously increases as $|\rho|$ increases. If $\lambda_1 = \lambda_2$, this error reduces to $2\rho\mu_I/3$, which has a maximum absolute value of $\mu_I/6$.

It is apparent from Table 1 and Equations (2.5) and (2.8) that the error in modeling system reliability and mean system life, based on independence, increases as

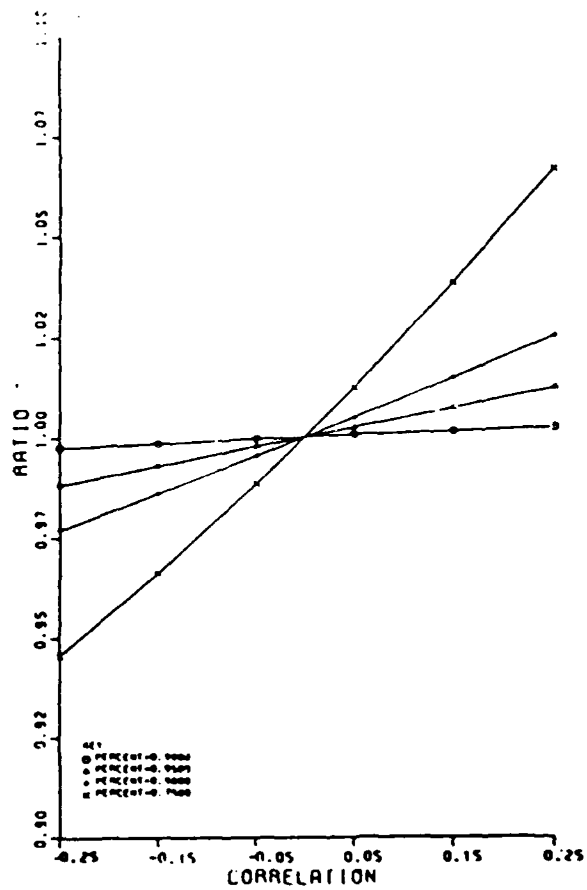


Figure 2. *Ratio of 100 pth Percentile Under Dependence and Independence Versus Correlation for* $\lambda_1 = 1, \lambda_2 = 1.5$.

$|\rho|$ increases and is a function of the relative sizes of λ_1 and λ_2. In particular, when the mean life of one component is substantially greater than the mean life of the second component, then the behavior of the system is well approximated by the behavior of the shorter-lived component acting alone. This can be seen in (2.4) and (2.7) by letting $\lambda_1 \to 0$. In this instance we also see, from (2.5) and (2.8), that the amount of error incurred by assuming independence is negligible.

3. ESTIMATING SYSTEM RELIABILITY FROM COMPONENT DATA

A common practice in predicting system reliability is to test each of the components independently and then to use the data to obtain confidence bounds on

Table 1. *Maximum Values of* $|\Delta(t)|$ *for* $\lambda_1 = 1$ *and Various Values of* λ_2

λ_2	Max $\lvert\Delta(t)\rvert$
2	.056
4	.041
8	.025
16	.014

system reliability. These bounds, obtained by Mann and Grubbs (1974), assume that the component lifetimes are exponential and that the components act independently when linked in series. In the bivariate case the bounds are computed as follows: For the jth component, suppose that n_j prototypes have been tested until $r_j\ (\le n_j)$ failures occur. Let Z_j be the total time on test for the jth component. Define

$$M^* = \sum (r_j - 1)/Z_j + \frac{\sum (r_j - 1)/Z_j^3}{\sum (r_j - 1)/Z_j^2}, \qquad (3.1)$$

and

$$V^* = \sum (r_j - 1)/Z_j^2 + \frac{\sum (r_j - 1)/Z_j^4}{\sum (r_j - 1)/Z_j^3}. \qquad (3.2)$$

An approximate γ-level lower confidence bound for system reliability at time t_m is

$$\exp\{-t_m M^*\{1 - V^*/(9M^{*2}) + \eta_\gamma (V^*)^{1/2}/(3M^*)\}^3\}, \qquad (3.3)$$

where η_γ is the 100γ percentile of a standard normal random variable.

When the system being evaluated has dependent components, these bounds may be misleading. The problem is that component data are independent, since the components are tested separately, but when they are put together into a system, some interdependence may develop. Of course, such dependence is not detectable in the absence of some system data, since the data on components we see are independent. To study the performance of the bound (3.3) when the correct system model is the Gumbel model (2.2), a simulation study was performed. For each simulated sample, n_j observations from exponential populations with mean $1/\lambda_j$, $j = 1, 2$, were simulated. The two samples were generated independently. The confidence bound (3.3) was obtained. This was then compared to the true system reliability at various ρ's obtained from (2.4). Ten thousand such bounds were simulated for each set of parameter values. The estimated coverage probabilities for the Mann-Grubbs bounds (i.e., the proportion of times that the Mann-Grubbs intervals assuming independence actually contained the true system reliability) for $n_1 = n_2 = 3$, 5, 10, $\lambda_1 = 1.0$, $\lambda_2 = 1.5$, at $t_m = .1$ are reported in Table 2. Here the true system reliability under dependence ranges from .7684 at $\rho = -.25$ to .7891 at $\rho = .25$, with a value of .7788 at $\rho = 0$.

The results in Table 2 show that at high negative correlations, the coverage probabilities are significantly lower than claimed under independence, and for a high positive correlation, the intervals are conservative. This trend becomes more exaggerated as n_1, n_2 increase because the bound approaches the reliability under independence. As seen in Section 2, the true reliability at t is an increasing function of ρ so that asymptotically coverage probabilities approach 0 (or 1) for $\rho < 0\ (>0)$. For sample sizes in the range of 3 to 10, the estimated coverage probabilities for $\rho < 0$ are statistically significantly lower than expected. On the practical side, however, they are not of sufficient magnitude to cause great concern, especially at $\gamma = .95$.

Table 2. *Estimated Coverage Probabilities for Mann-Grubbs Bounds* ($\lambda_1 = 1.0$, $\lambda_2 = 1.5$)

			Correlation						
n_1	n_2	γ	−.25	−.15	−.05	0	.05	.15	.25
3	3	.95	93.41*	94.11*	94.74	95.05	95.27	95.80*	96.22*
3	3	.90	87.40*	88.42*	89.32*	89.78	90.20	91.18*	92.15*
3	3	.75	71.03*	72.53*	74.13*	74.88	76.58	77.34*	78.81*
5	5	.95	93.19*	94.04*	94.90	95.26	95.62*	96.17*	96.81*
5	5	.90	87.12*	88.48*	89.85	90.39	91.10	92.13*	93.14*
5	5	.75	69.68*	72.02*	74.10*	75.13	76.14*	78.32*	80.03*
10	10	.95	92.03*	93.42*	94.58	95.08	95.51*	96.42*	97.14*
10	10	.90	85.93*	87.70*	89.34*	90.21	91.05*	92.56*	93.93*
10	10	.75	67.77*	70.90*	74.12*	75.63	77.05*	79.87*	82.56*

* At least two standard errors above specified level

* At least two standard errors below specified level

NOTE: Standard errors of the above estimates are approximately .2 for the .95 level, .3 for the .90 level, and .4 for the .75 level.

4. PARAMETRIC ESTIMATION OF COMPONENT PARAMETERS

In this section we are interested in examining how the independence assumption affects the magnitude of the estimation error in estimating component mean life from data collected on series systems. That is, for each system tested, we observe its failure time and an indicator variable that tells us which component caused the system to fail. We are interested in how varying degrees of dependence affect the bias and mean squared error (MSE) of the maximum likelihood estimator of component mean life obtained by assuming independent component lifetimes.

We assume that the two components' survival functions are $\bar{F}_i(t) = \exp(-\lambda_i t)$, $i = 1, 2$, and a life test is conducted by putting n systems on test. We observe n_i systems failing because of failure of the ith component, $i = 1, 2$. Let T denote the sum of all n failure times. From Moeschberger and David (1971), the maximum likelihood estimator of λ_i, assuming independence, is

$$\hat{\lambda}_i = n_i/T, \qquad i = 1, 2,$$

so the estimator of component mean life, $\mu_i = \lambda_i^{-1}$, is

$$\hat{\mu}_i = T/n_i \quad \text{if} \quad n_i > 0 \qquad (4.1)$$

Now suppose that we are in fact sampling from the Gumbel distribution (2.3). For this model, component mean life is the same as in the independent case. The random variables (n_i, T) are independent (the conditional distribution of T given n_i is free of n_i), and n_i is binomial with parameters n and $p_i = P(\min(X_1, X_2) = X_i)$. For this model,

$$p_1 = P(X_1 < X_2)$$
$$= \lambda_1\left\{\frac{1}{\lambda_1 + \lambda_2} + \frac{4\rho(\lambda_1 - \lambda_2)\lambda_2}{(\lambda_1 + \lambda_2)(2\lambda_1 + \lambda_2)(\lambda_1 + 2\lambda_2)}\right\}, \tag{4.2}$$

with $p_2 = 1 - p_1$. From Mendenhall and Lehman (1960), approximations to the moments of $1/n_i$, conditional on $n_i > 0$, are

$$E(1/n_i \mid n_i > 0) = \frac{n-2}{n(a-1)} \tag{4.3}$$

and

$$E(1/n_i^2 \mid n_i > 0) = \frac{(n-2)(n-3)}{n^2(a-1)(a-2)}, \tag{4.4}$$

where $a = (n-1)p_i$. The expected value of T is given by $n\mu_D$, where μ_D is given by (2.7), and

$$E(T^2) = n\left[\frac{2+10\rho}{(\lambda_1+\lambda_2)^2} - 8\rho\left(\frac{1}{(2\lambda_1+\lambda_2)^2} + \frac{1}{(\lambda_1+2\lambda_2)^2}\right)\right] + n(n-1)\mu_D^2. \tag{4.5}$$

Thus, the bias and MSE of $\hat{\mu}_i$, conditional on $n_i > 0$, under this model are

$$B(\hat{\mu}_i) = E(\hat{\mu}_i - \mu_i) = \frac{(n-2)\mu_D}{[(n-1)(p_i-1)]} - \mu_i, \tag{4.6}$$

and

$$\text{MSE}(\hat{\mu}_i) = E(T^2)E(1/n_i^2 \mid n_i > 0) - \frac{2\mu_i(n-2)\mu_D}{[(n-1)p_i - 1]} + \mu_i^2, \tag{4.7}$$

for $i = 1, 2$.

We note that for large samples,

$$\lim_{n\to\infty} B(\hat{\mu}_i) = \frac{\mu_D}{p_i} - \mu_i \tag{4.8}$$

$$\lim_{n\to\infty} \text{MSE}(\hat{\mu}_i) = \left(\lim_{n\to\infty} B(\hat{\mu}_i)\right)^2 \tag{4.9}$$

for $i = 1, 2$. For $\lambda_1 = \lambda_2$ from (4.6), we see that

$$B(\hat{\mu}_1) = \frac{1 + 2(n-2)\rho/3}{(n-3)}\mu_1$$
$$= \frac{\mu_1}{n-3} + \frac{2(n-2)\rho\mu_1}{3(n-3)}. \tag{4.10}$$

A similar expression holds for $B(\hat{\mu}_2)$. Note that (4.10) consists of two terms. The first term, reflecting sampling error, is positive for all n and dominates the bias expression for small n. The second term, reflecting modeling error, takes on the same sign as the correlation and dominates for large n, approaching the limit of $2\rho\mu_1/3$.

When $\lambda_1 = \lambda_2$,

$$\text{MSE}(\hat{\mu}_1) = \frac{2\mu_1^2(n^2-2n+3)}{n(n-5)(n-3)} + \frac{2\mu_1^2(n-2)}{9n(n-5)(n-3)} \times \{(19n-21)\rho + 2(n-3)(n-1)\rho^2\}. \tag{4.11}$$

As in the bias expression, the MSE reflects a sampling error term and a modeling error term. The modeling error is a quadratic function of ρ for fixed n. For $n > 5$, this error is increasing in ρ for

$$\rho > -\frac{1}{4}\frac{(19n-21)}{(n-3)(n-1)}$$

and decreasing in ρ for

$$\rho < -\frac{1}{4}\frac{(19n-21)}{(n-2)(n-1)}$$

For sample sizes between 5 and 21, the modeling error, and hence the MSE, is a strictly increasing function for all $\rho \in [-\frac{1}{4}, \frac{1}{4}]$. For $n > 21$, the minimum MSE is achieved at $\rho < 0$. As n approaches ∞, the value at which the smallest MSE occurs tends to 0.

For unequal component means a similar result holds. Figure 3 shows the bias as a function of ρ for

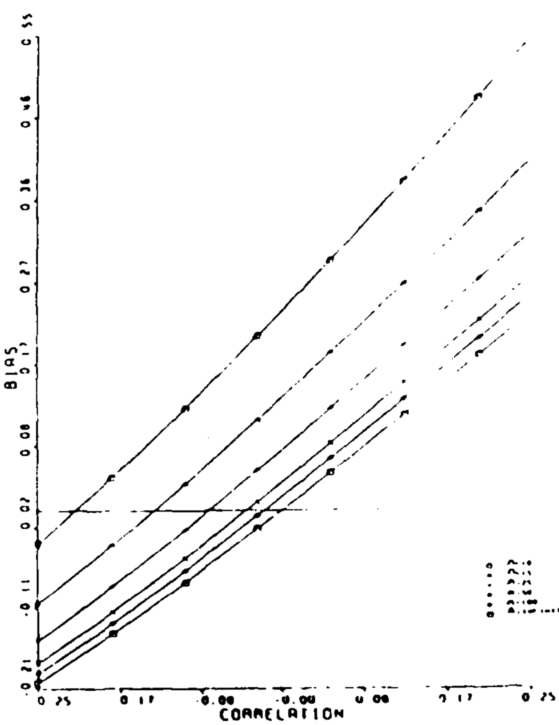


Figure 3. *Bias of $\hat{\mu}_1$ Under Gumbel's Model for $\lambda_1 = 1$, $\lambda_2 = 1.5$.*

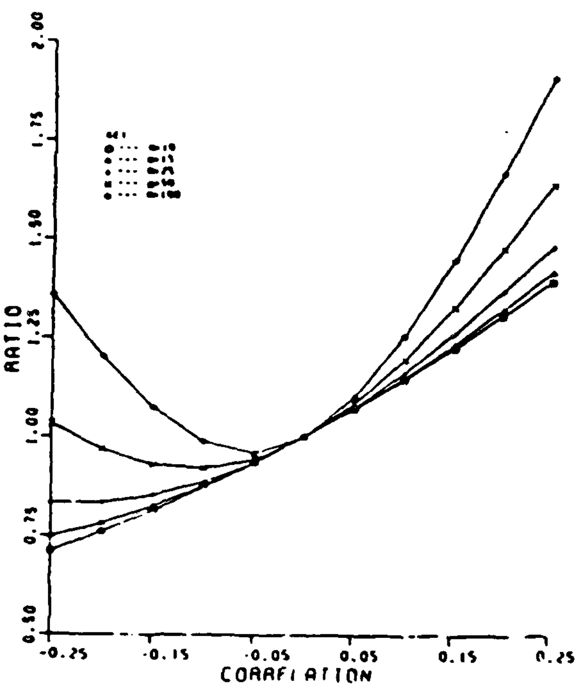


Figure 4. *Ratio of* $\sqrt{MSE\ (\hat{\mu}_1 \mid \rho)/MSE\ (\hat{\mu}_1 \mid \rho = 0)}$ *for Various Sample Sizes n and for* $\lambda_1 = 1, \lambda_2 = 1.5$.

various sample sizes when $\lambda_1 = 1.0$ and $\lambda_2 = 1.5$. Figure 4 depicts the ratio

$$\sqrt{MSE\ (\hat{\mu}_1 \mid \rho)/MSE\ (\hat{\mu}_1 \mid \rho = 0)}$$

as a function of ρ for various sample sizes when $\lambda_1 = 1, \lambda_2 = 1.5$.

5. BIAS OF THE PRODUCT LIMIT ESTIMATOR

A second approach to the problem of estimating component parameters is via the nonparametric estimator proposed by Kaplan and Meier (1958). Investigators who routinely use nonparametric techniques may take this approach in hopes of obtaining estimators that are robust with respect to the assumption of exponentiality. The purpose of this section is to show that such estimators are not necessarily robust with respect to the assumption of independence when the marginals are, in fact, exponential.

The product limit estimator, assuming independent risks, is constructed as follows. As before, suppose n systems are put on test at time 0 and n_i systems fail owing to failure of component i. Let $X_{i(1)}, \ldots, X_{i(n_i)}$ denote the ordered times at which these n_i events occur, and let $r_{i1}, \ldots, r_{in_i}$ be the ranks of those ordered survival times among all n ordered lifetimes. The component reliability for the ith component at time x may now be estimated by the product of the individual conditional survival probabilities, namely, by

$$\hat{F}_i(x) = 1 \quad \text{if} \quad x < x_{i(1)}$$

$$= \prod_{j=1}^{j(i,x)} \frac{n - r_{ij}}{n - r_{ij} + 1}, \qquad x > x_{i(1)},$$

where $j(i, x)$ is the largest value of j for which $x_{i(j)} < x$. A special note is needed to cover the case in which $x_{i(n_i)}$ is not the largest observed death. To avoid this problem, we shall define $\hat{F}_i(x) = 0$ for x greater than the largest observed failure time.

If the component lifetimes in fact follow the Gumbel bivariate exponential, we can see that the Kaplan-Meier estimator is not consistent. For $i = 1$, the Kaplan-Meier estimator is not estimating $F_1(t)$, but, rather, another survival function, $H_1(t)$, given by

$$H_1(t) = \exp\left\{-\lambda_1 \int_0^t \frac{[1 + 4\rho(1 - e^{-\lambda_2 u})(1 - 2e^{-\lambda_1 u})]}{[1 + 4\rho(1 - e^{-\lambda_1 u})(1 - e^{-\lambda_2 u})]}\, du\right\}, \quad t > 0. \tag{5.1}$$

Note that if $\lambda_1 = \lambda_2 = \phi$, (5.1) is simplified to

$$H_1(t) = e^{-\phi t}[1 + 4\rho(1 - e^{-\phi t})^2]^{1/2}, \tag{5.2}$$

which is increasing in ρ. Similarly, $\hat{F}_2(t)$ is actually estimating $H_2(t)$, which is defined analogously.

Measures of the error in estimating $F_i(t)$ by $\hat{F}_i(t)$ are

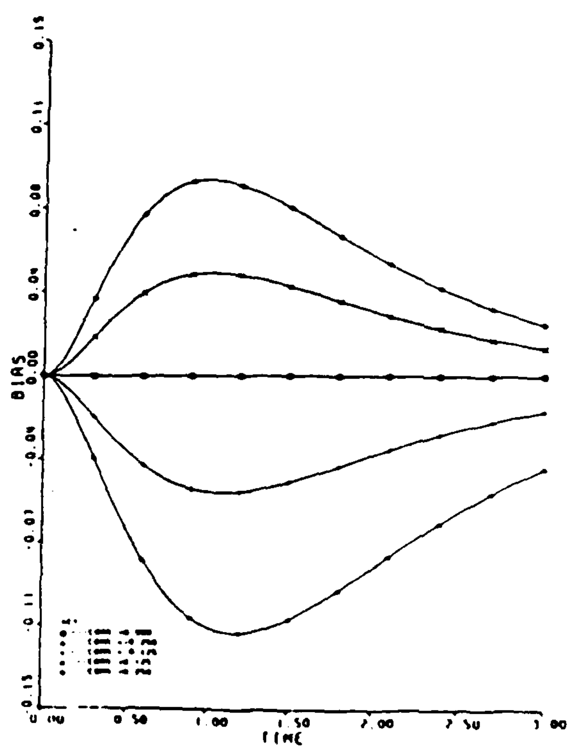


Figure 5. *Bias of Kaplan-Meier Estimate,* $\hat{F}_1(t)$, $\lambda_1 = 1, \lambda_2 = 1.5$.

the bias and MSE of $\bar{F}_i(t)$ computed under the dependence model. Under this model, the Kaplan-Meier estimator is equivalent to the estimator one would obtain based on n observations from an independent system with component survival distributions $\bar{H}_i$ given by (5.1) or, if $\lambda_1 = \lambda_2$, by (5.2). Hence from Kaplan and Meier (1958), the variance of $\hat{\bar{F}}_i(t)$ is given by

$$V(\hat{\bar{F}}_i(t)) = \bar{H}_i(t)^2 \int_0^t \frac{|d\bar{H}_i(u)|}{n\bar{H}_i(u)^2}. \tag{5.3}$$

Thus from (5.1) and (5.2), the bias and MSE of $\hat{\bar{F}}_i(t)$ are

$$B(\hat{\bar{F}}_i(t)) = \bar{H}_i(t) - \bar{F}_i(t), \qquad t \geq 0, \tag{5.4}$$

and

$$\text{MSE}\,(\hat{\bar{F}}_i(t)) = (\bar{H}_i(t) - \bar{F}_i(t))^2 + \bar{H}_i(t)^2 \int_0^t \frac{|d\bar{H}_i(u)|}{n\bar{H}_i(u)^2}, \qquad t > 0. \tag{5.5}$$

The estimator is not consistent, since $B(\hat{\bar{F}}_i(t))$ is independent of n and not necessarily zero. Also, MSE $(\hat{\bar{F}}_i(t))$ consists of a factor that depends only on the model error and is free of sample size and of a term that tends to 0 as n tends to infinity.

Note that in the case of equal component lifetime distributions, $\lambda_1 = \lambda_2 = \phi$, the bias determined from (5.2) and (5.4) simplifies to

$$B(\hat{\bar{F}}_i(t)) = e^{-\phi t}\{[1 + 4(1 - e^{-\phi t})^2]^{1/2} - 1\}. \tag{5.6}$$

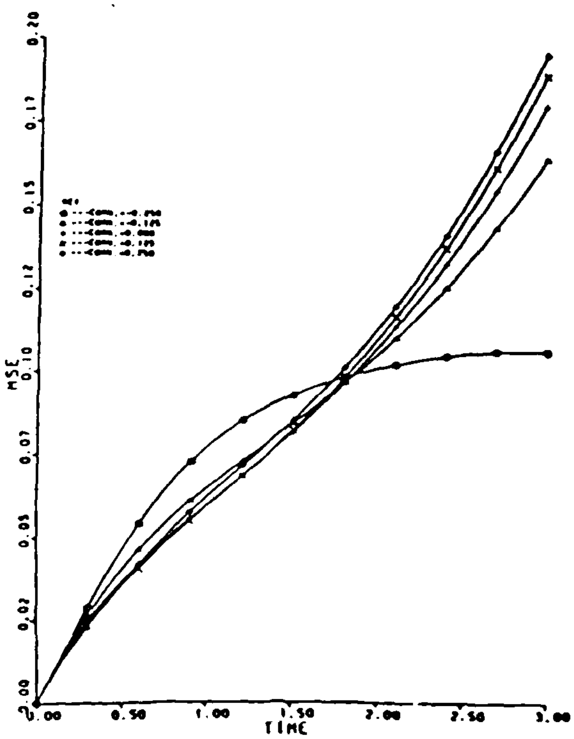


Figure 6. *MSE of Kaplan–Meier Estimate, $\bar{F}_1(t)$, $\lambda_1 = 1, \lambda_2 = 1.5, n = 10$.*

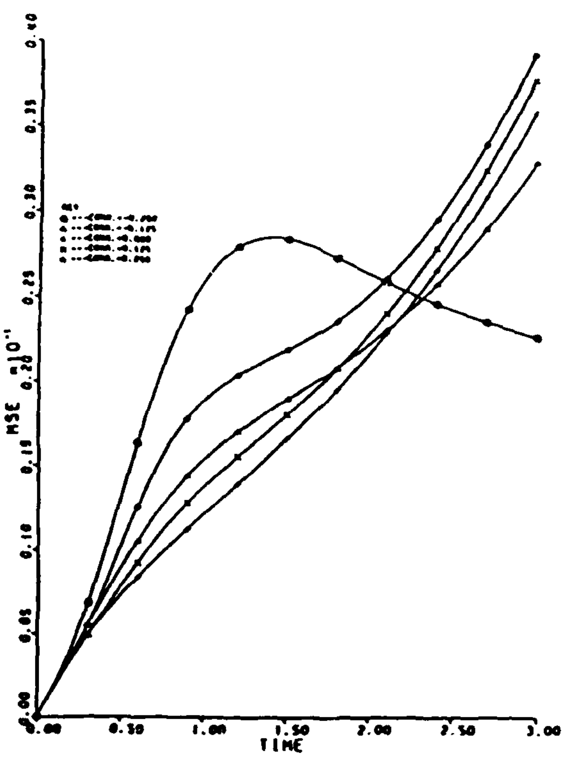


Figure 7. *MSE of Kaplan–Meier Estimate, $\bar{F}_1(t)$, $\lambda_1 = 1, \lambda_2 = 1.5, n = 50$.*

In the general case, the integral in (5.1) needs to be evaluated numerically. The bias of the Kaplan-Meier estimator was calculated for various values of λ_i and ρ. A representative plot of the bias appears in Figure 5, where $\lambda_1 = 1$, $\lambda_2 = 1.5$, and $|\rho| = 0, .125, .250$. It is apparent that the bias is largest for values of t in the neighborhood of an interval that captures the mean component lifetimes. The absolute value of the bias ranges from 0 to .11 in this example.

MSE $(\hat{\bar{F}}_i(t))$ was calculated for various values of λ_i, n, and ρ. Its magnitude is typified in Figures 6 and 7, where $\lambda_1 = 1$, $\lambda_2 = 1.5$, and $n = 10, 50$, respectively. For $\lambda_1 = 1$, $\lambda_2 = 1.5$, and $n = \infty$, MSE $(\hat{\bar{F}}(t))$ may be obtained by squaring (5.4) or by squaring the ordinate values in Figure 5. The MSE of the Kaplan-Meier estimator may be quite large for small sample size n and moderately large for "large" ρ, the former being a more crucial factor than the latter.

6. SUMMARY

The results presented here show that for the Gumbel model, one may be misled by falsely assuming independence of component lifetimes in a series system. In modeling system reliability based on complete information about two marginal component life distributions, effects of erroneously assuming independence of component lifetimes is most pronounced for system reliabilities smaller than .75. For system reliabilities larger than .90, this effect is too small to be of practical interest. The effects of a departure from independence on the Mann-Grubbs bounds for small sample sizes seems to be negligible for confidence levels greater than .90. But for either large sample sizes or smaller confidence levels, one may be appreciably misled.

For the dual problem of estimating component reliability based on data from a series system, it appears that departures from independence are of a greater consequence. Both parametric and nonparametric estimators of relevant component parameters are inconsistent. Although under independence, the bias of the estimators of interest clouds the issues, it is clear that for larger negative correlations these estimators tend to underestimate the parameter, whereas for large positive correlations, the reverse is true.

ACKNOWLEDGMENTS

The authors would like to thank the editor and the referees for their constructive comments.

This work was supported by the Air Force Office of Scientific Research under Contract AFOSR-82-0307 and the Graduate School of the Ohio State University.

[*Received June 1982. Revised December 1983.*]

REFERENCES

BARLOW, R. E., and PROSCHAN, F. (1975), *Statistical Theory of Reliability and Life Testing*, New York: Holt, Rinehart, and Winston.

BASU, A. P. (1981), "Identifiability Problems in the Theory of Competing and Complementary Risks—A Survey," in *Statistical Distribution in Scientific Work*, eds. Taillie, Patil, and Baldessari, Dorerecht, Holland: Reidel Publishing Co., 335–348.

BIRNBAUM, Z. W. (1979), *On the Mathematics of Competing Risks*, U.S. Dept. of HEW Publication 79-1351.

BOARDMAN, T. J., and KENDALL, P. J. (1970), "Estimation in Compound Exponential Failure Models," *Technometrics*, 12, 891–900.

CHAO, ANNE (1981), "Approximate Mean Squared Errors of Estimators of Reliability for *k*-out-of-*m* Systems in the Independent Exponential Case," *Journal of the American Statistical Association*, 76, 720–724.

EASTERLING, R. G. (1980), Review of *On the Mathematics of Competing Risks*, by Z. W. Birnbaum, *Technometrics*, 22, 131–132.

GUMBEL, E. J. (1960), "Bivariate Exponential Distributions," *Journal of the American Statistical Association*, 55, 698–707.

KALBFLEISCH, J. D., and PRENTICE, R. L. (1980), *The Statistical Analysis of Failure Time Data*, New York: John Wiley.

KAPLAN, E. L., and MEIER, P. (1958), "Nonparametric Estimation From Incomplete Observations," *Journal of the American Statistical Association*, 53, 457–481.

LAGAKOS, S. W. (1979), "General Right Censoring and Its Impact on the Analysis of Survival Data," *Biometrics*, 35, 139–156.

LAGAKOS, S. W., and WILLIAMS, J. S. (1978), "Models for Censored Survival Analysis: A Cone Class of Variable-Sum Models," *Biometrika*, 65, 181–189.

LAWLESS, J. F. (1982), *Statistical Models and Methods for Lifetime Data*, New York: John Wiley.

MANN, N. R. (1974), "Simplified Expressions for Obtaining Approximately Optimum System-Reliability Confidence Bounds From Exponential Subsystem Data," *Journal of the American Statistical Association*, 69, 492–495.

MANN, N. R., and GRUBBS, F. E. (1974), "Approximately Optimum Confidence Bounds for System Reliability Based on Component Test Data," *Technometrics*, 16, 335–347.

MANN, N. R., SCHAFER, R. E., and SINGPURWALLA, N. D. (1974), *Methods for the Statistical Analysis of Reliability and Life Data*, New York: John Wiley.

MENDENHALL, W., and LEHMAN, E. H., JR. (1960), "An Approximation to the Negative Moments of the Positive Binomial Useful in Life Testing," *Technometrics*, 2, 227–242.

MIYAMURA, T. (1982), "Estimating Component Failure Rates From Combined Component and Systems Data: Exponentially Distributed Component Lifetimes," *Technometrics*, 24, 313–318.

MOESCHBERGER, M. L., and DAVID, H. A. (1971), "Life Tests Under Competing Causes of Failure and the Theory of Competing Risks," *Biometrics*, 27, 909–933.

PETERSON, A. V., JR. (1976), "Bounds for a Joint Distribution Function With Fixed Sub-Distribution Functions: Application to Competing Risks," *Proceedings of the National Academy of Sciences*, 73, 11–13.

TSIATIS, A. (1975), "A Nonidentifiability Aspect of the Problem of Competing Risks," *Proceedings of the National Academy of Sciences*, 72, 20–22.

WILLIAMS, J. S., and LAGAKOS, S. W. (1977), "Models for Censored Survival Analysis: Constant-Sum and Variable-Sum Models," *Biometrika*, 64, 215–224.

APPENDIX B

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY
NA

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE
NA

3. DISTRIBUTION / AVAILABILITY OF REPORT
Approved for public release; distribution unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S)

5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION
The Ohio State University

6b. OFFICE SYMBOL (If applicable)

7a. NAME OF MONITORING ORGANIZATION
AFOSR/NM

6c. ADDRESS (City, State, and ZIP Code)
1314 Kinnear Road
Columbus, Ohio 43212

7b. ADDRESS (City, State, and ZIP Code)
AFOSR/NM
Bldg 410
Bolling AFB DC 20332-6448

8a. NAME OF FUNDING / SPONSORING ORGANIZATION
AFOSR

8b. OFFICE SYMBOL (If applicable)
NM

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
AFOSR-82-0307

8c. ADDRESS (City, State, and ZIP Code)
AFOSR/NM
Bldg 410
Bolling AFB DC 20332-6448

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
61102F	2304		

11. TITLE (Include Security Classification) Asymptotic Bias of the Product Limit Estimator under Dependent Competing risks

12. PERSONAL AUTHOR(S)
John P. Klein and M.L. Moeschberger

13a. TYPE OF REPORT
Final

13b. TIME COVERED
FROM 9-1-82 TO 12-31-87

14. DATE OF REPORT (Year, Month, Day)
May 31, 1988

15. PAGE COUNT
7

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP
XXXX	XXXXXXXXXX	XXX

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Bias of Kaplan-Meier, product-limit estimator, dependent competing risks

19. ABSTRACT (Continue on reverse if necessary and identify by block number)
A common assumption made in analyzing competing risk experiments is that the risks are stochastically independent. Under that assumption the product limit estimator is a consistent estimator of the marginal survival function. We show that when the risks are not independent the product limit estimator converges, with probability one, to a survival function which may not be the same as the marginal survival function of interest.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT
☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL
Maj. Brian Woodruff

22b. TELEPHONE (Include Area Code)
(202) 767-5027

22c. OFFICE SYMBOL
NM

DD FORM 1473, 84 MAR
83 APR edition may be used until exhausted.
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

Reprinted from
IAPQR TRANSACTIONS
Journal of the Indian Association
for Productivity, Quality & Reliability
Vol. 9, *No.* 1, 1984

IAPQR Transactions
Vol. 9, No. 1, 1984

Asymptotic Bias of the Product Limit Estimator under Dependent Competing Risks

JOHN P. KLEIN
Ohio State University
AND
M. L. MOESCHBERGER
Ohio State University

ABSTRACT

A common assumption made in analyzing competing risk experiments is that the risks are stochastically independent. Under that assumption the product limit estimator is a consistent estimator of the marginal survival function. We show that when the risks are not independent the product limit estimator converges, with probability one, to a survival function which may not be the same as the marginal survival function of interest.

1. Introduction

Competing risks arise in a wide range of life testing problems. Typical areas of application are the study of series systems in the engineering sciences and biological systems in the medical sciences. An important area of application is the analysis of censored data where some systems or individuals are lost or withdrawn from a study prior to observing the endpoint of interest. Competing risks are often modeled by a vector $\underset{\sim}{T} = (T_1, \ldots, T_p)$ of nonnegative random variables representing the potential times to failure from each of the p causes. We cannot observe $\underset{\sim}{T}$ directly but instead we see the system failure time $Y = \min(T_i, i = 1, \ldots, p)$ and the failure pattern $\xi(\underset{\sim}{T}) = I$ such that $Y = T_i$ for $i \in I$ and $Y < T_i$ $i \notin I$, where $I \in \mathcal{I}$ the set of all subsets of $1, \ldots, p$. Based on this information we wish to estimate the marginal survival probabilities

$$S_I(t) = P(\min(T_i, i \in I) > t), \quad t > 0, \ I \in I.$$

A common assumption made in analyzing competing risk experiments is that the T_i's are independent random variables. Such an assumption is not testable due to the identifiability dilemma (see Basu (1981)). Under an assumption of independent risks a consistent estimator of $S_I(t)$ is the Kaplan-Meier (1958) product limit estimator. In Section 2 we show that, if the risks are dependent, then the product limit estimator may be an inconsistent estimator of $S_I(t)$. The quantity to which this estimator converges is obtained so that one may investigate the estimator's robustness to departures from independence. In Section 3 we illustrate such robustness considerations for some well-known bivariate exponential distributions.

2. Inconsistency of the Product Limit Estimator

The Kaplan-Meier product limit estimator is constructed as follows. Let $0 = Y_{(0)} \leqslant Y_{(1)} \leqslant \ldots \leqslant Y_{(n)}$ denote the ordered system failure times of n systems put on test. The product-limit estimator of $S_I(t)$ is

$$\hat{S}_I(t) = \Pi_i [(n - i) / (n - i + 1)] \tag{2.1}$$

where the product is over the ranks i of those ordered observations $Y_{(i)}$, $1 \leqslant i \leqslant n$, such that $Y_{(i)} \leqslant t$ and $Y_{(i)}$ corresponds to a death from the simultaneous cause (s) $j \in J$, $J \cap I \neq \phi$. $\hat{S}_I(t)$ is undefined for $t > Y_{(n)}$ if the largest failure time corresponds to causes in J where $J \cap I = \phi$.

If the assumption of independence is correct and the crude probability functions defined by $F(t, I) = P(Y > t, \xi(\underset{\sim}{T}) = I)$ have no common discontinuities then Langberg, Proschan and Quinzi [LPQ (1981)] have shown that the product limit estimator is consistent. They also show that if the $F(t, I)$'s have no common discontinuities then for a very particular form of dependence structure the product limit estimator is consistent. We note in the following theorem, that their results can be used to study the robustness of the product limit estimator to departures from independence and that, in general, if the risks are dependent then the product limit estimator (2.1) is inconsistent.

Theorem 1.

Let $\underset{\sim}{T} = (T_1, \ldots, T_p)$ be a vector of non-negative random variables with system life $Y = \min(T_1, \ldots, T_p)$ and failure pattern

$$\xi(\underset{\sim}{T}) = \mathrm{I}, \text{ if } Y = T_i,\ i \in \mathrm{I} \text{ and } Y < T_j,\ j \notin \mathrm{I} \tag{2.2}$$

$$= \phi, \quad \text{otherwise.}$$

Define $\bar{F}(t) = P(Y > t)$, $F(t, \mathrm{I}) = P(Y < t, \xi(\underset{\sim}{T}) = \mathrm{I})$, $\mathrm{I} \in I$, $t > 0$, and let $\alpha(\bar{F}) = \{t : \bar{F}(t) > 0\}$ be the support of $\bar{F}$. For $\mathrm{I} \in I$ define $I_{\mathrm{I}} = \{J \in I : J \cap \mathrm{I} \neq \phi\}$. Based on a random sample of size n let $\hat{S}_{\mathrm{I},n}(t)$ be the product limit estimator (2, 1). If the functions $F(\cdot, \mathrm{I})$ have no common discontinuities on $[0, \alpha(\bar{F})]$ then

$$\hat{S}_{\mathrm{I},n}(t) \to \prod_{J \in I_{\mathrm{I}}} \bar{G}_{\mathrm{J}}(t) \quad \text{a. s.} \tag{2.3}$$

where

$$\bar{G}_{\mathrm{J}}(t) = \prod_{a \leq t} [\bar{F}(a) / \bar{F}(a^-)] \exp[-\int_0^t (dF^c(\cdot, \mathrm{J}) / \bar{F})],$$

$$0 \leq t \leq \alpha(\bar{F}). \tag{2.4}$$

where the product is over the set of discontinuities of $F(\cdot, \mathrm{J})$ and $F^c(\cdot, \mathrm{J})$ is the continuous part of $F(\cdot, \mathrm{J})$.

Proof

The proof follows directly by applying the results of Langberg, Proschan and Quinzi [LPQ (1978)] and LPQ (1981). By Theorem 4.1 of LPQ (1978) $\underset{\sim}{T} =_{LP} \underset{\sim}{H}$ where H is a vector of $(2^p - 1)$ independent components indexed lexicographically by $\mathrm{I} \in I$ with $P(H_{\mathrm{I}} > t) = \bar{G}_{\mathrm{I}}(t)$ given by (2.4).

($T =$ LPH if $P(Y > t, \xi(\underset{\sim}{T}) = \mathrm{I}) = P(\min_j H_j > t, \xi(\underset{\sim}{H}) = \mathrm{I})$).

Let $\underset{\sim}{T_i} = (T_{1i}, \ldots, T_{pi})$. $i = 1, \ldots, n$ be independent and identically distributed as $\underset{\sim}{T}$. Replace $\bar{F}$ and $F(\cdot, \mathrm{J})$ in (2.4) by their empirical counterparts $\hat{\bar{F}}_n(t) = \sum_{i=1}^{n} \chi\{Y_i > t\} / n$ and $\hat{F}_n(t, \mathrm{J}) = \sum_{i=1}^{n} \chi\{Y_i \leq t,$

$\xi(\underset{\sim}{T_i}) = J\} / n$, to obtain $\hat{\bar{G}}_{I,n}(t)$. Here $\chi(A)$ is the indicator function of the set A.

By Theorem 4.7 of LPQ [1981] $\hat{\bar{G}}_{I,n}(t) \to \bar{G}_J(t)$ a. s. for $J \in I$. Routine algebraic manipulation shows that $\hat{S}_{I,n}(t) = \prod_{J \in I_I} \hat{\bar{G}}_{J,n}(t)$ so the result now follows $\square$.

In general, as seen in the examples in the following section, $\prod_{J \in I_I} \bar{G}_J(t) \neq S_I(t)$ so that an investigator may be seriously misled by incorrectly assuming that the component lifetimes are independent. This has been noticed by Fisher and Kanarek (1974) in the problem of analyzing clinical trials with censored data. Theorem 1 allows an investigator to quantify the effects of the independence assumption by computing the right hand side of (2.3) for some plausible dependent models.

LPQ (1981) have shown that for a special type of dependence the estimator $\hat{S}_{I,n}(t)$ is consistent. We state their result, without proof, as a corollary.

Corollary 1.

Assume that the conditions of Theorem 1 hold. Then $\hat{S}_{I,n}(t) \to S_I(t)$ a. s. if and only if the following two conditions hold.

(i) $S_I(a) / S_I(a^-) = \bar{F}(a) / \bar{F}(a^-)$, a discontinuity point of $\Sigma F(\ , J)$ where the sum is over $J \in I_I$

$= 1$, otherwise.

(ii) $P(\min(T_i, i \in I') \geq t \mid \min(T_i, i \in I) = t)$

$= P(\min(T_i, i \in I') > t \mid \min(T_i, i \in I) > t)$

where I′ is the complement of $I \in I$.

3. EXAMPLES

In this section we present some representative examples of the use of Theorem 1 in determining the effects, on estimating marginal survival, of the independence assumption for some bivariate exponential life distributions. Let (T_1, T_2) denote the time to failure from components 1 and 2, respectively, in a series system. Let $\bar{F}(t_1, t_2)$ be the joint survival function of (T_1, T_2) and $S_i(t) = P(T_i > t)$, $i = 1, 2$ the marginal survival functions. Let $\hat{S}_i(t)$ be the estimator (2.1) of $S_i(t)$, and let (t), $J = \{1\}, \{2\}, \{1, 2\}$ be given by (2.4). Then $\hat{S}_i(t) \to \bar{G}_{\{i\}}(t)\, \bar{G}_{\{1,2\}}(t)$ a. s. by Theorem 1, if the functions $F(t, I)$ have no common discontinuities. Note that if $P(T_1 = T_2) = 0$ then $\bar{G}_{\{1,2\}}(t) = 1$ for all t. We now give some examples.

Example 1. (Block and Basu (1974))

$$\text{Let } \bar{F}(t_1, t_2) = [\lambda / (\lambda_1 + \lambda_2)] \exp(-\lambda_1 t_1 - \lambda_2 t_2 - \lambda_{12} \max(t_1, t_2)] - [\lambda_{12} / (\lambda_1 + \lambda_2)] \exp(-\lambda \max(t_1, t_2)), \quad \text{for } t_1, t_2 > 0,$$

$$\lambda_1, \lambda_2 > 0, \ \lambda_{12} \geq 0, \ \lambda = \lambda_1 + \lambda_2 + \lambda_{12}.$$

$$\text{Here } S_i(t) = \frac{\lambda}{(\lambda_1 + \lambda_2)} \exp(-(\lambda_i + \lambda_{12})\, t) - \frac{\lambda_{12}}{(\lambda_1 + \lambda_2)} \exp(-\lambda t), \quad t > 0, \text{ but}$$

by theorem 1

$$\hat{S}_i(t) \to \exp\left(-\frac{\lambda_i \lambda}{(\lambda_1 + \lambda_2)}\, t\right), \text{ a. s. } t > 0, \ i = 1, 2.$$

Example 2. Gumbel (1960)

$$\text{Let } \bar{F}(t_1, t_2) = \exp(-\lambda_1 t_1 - \lambda_2 t_2 - \lambda_{12} t_1 t_2)\ t_1, \ t_2 > 0, \quad \lambda_1 + \lambda_2 > 0, \ \lambda_{12} \geq 0.$$

Here $S_i(t) = \exp(-\lambda_i t)$, $t \geq 0$ but by Theorem 1

$$\hat{S}_i(t) \to \exp(-\lambda_i t - \lambda_{12} t^2 / 2), \text{ a. s. for } t \geq 0, \ i = 1, 2.$$

Example 3. Gumbel (1960)

$$\text{Let } \bar{F}(t_1, t_2) = \exp(-\lambda_1 t_1 - \lambda_2 t_2) [1 + \lambda_{12} - \lambda_{12} (\exp(-\lambda_1 t_1) + \exp(-\lambda_2 t_2))] + \lambda_{12} \exp(-\lambda_1 t_1 - \lambda_2 t_2)], \quad t_1, t_2 > 0, \ \lambda_1, \lambda_2 > 0, \ \lambda_{12} \geq 0.$$

Here $S_i(t) = \exp(-\lambda_i t)$ but by Theorem 1

$$\hat{S}_i(t) \rightarrow \exp\left\{-\lambda_i \int_0^t \frac{[1 + \lambda_{12}(1 - e^{-\lambda_j t})(1 - 2e^{-\lambda_i t})]}{[1 + \lambda_{12}(1 - e^{-\lambda_i t})(1 - e^{-\lambda_j t})]}\, dt\right.$$

where j is the compliment of i in $\{1, 2\}$.

Example 4. Marshall-Olkin (1967)

Let $\bar{F}(t_1, t_2) = \exp(-\lambda_1 t_1 - \lambda_2 t_2 - \lambda_{12} \max(t_1, t_2))$

$t_1, t_2 > 0,\ \lambda_1, \lambda_2 \geqslant 0,\ \lambda_{12} \geqslant 0.$

Here $S_i(t) = \exp(-(\lambda_i + \lambda_{12})\, t)$. In this case the conditions of Corollary 1 are met so $\hat{S}_i(t) \rightarrow S_i(t)$ a. s.

ACKNOWLEDGEMENT

This research was supported by Contract AFOSR-82-0307 for the Air Force Office of Scientific Research.

REFERENCES

1. Basu, A. P. (1981) : Identifiability problems in the theory of competing and complimentary risks—a survey, in *Statistical Distributions in Scientific Work*, eds. Taillies, Patil and Baldesaari, Dorerecht, Holland : Reidel Publishing Company, 335-348.

2. Block, H. W. and Basu, A. P. (1974) : A continuous bivariate exponential extension, *Jour. Amer. Stat. Assoc*, 69, 1031-1037.

3. Fisher, L. and Kanarek, P. (1974) Presenting censored survival data when censoring and survival times may not be independent, in *Reliability and Biometry : Statistical Analysis of Lifelength*, Eds. Proschan and Serfling, Philadelphia, Society for Industrial and Applied Mathematics, 303-326.

APPENDIX C

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY
NA

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE
NA

3. DISTRIBUTION/AVAILABILITY OF REPORT
Approved for Public Release; Distribution Unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S)

5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION
The Ohio State University

6b. OFFICE SYMBOL *(If applicable)*

7a. NAME OF MONITORING ORGANIZATION
AFOSR/NM

6c. ADDRESS *(City, State and ZIP Code)*
1314 Kinnear Road
Columbus, Ohio 43212

7b. ADDRESS *(City, State and ZIP Code)*
Bldg. 410
Bolling AFB, DC 20332-6448

8a. NAME OF FUNDING/SPONSORING ORGANIZATION
AFOSR

8b. OFFICE SYMBOL *(If applicable)*
NM

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
AFOSR-82-0307

8c. ADDRESS *(City, State and ZIP Code)*
Bldg. 410
Bolling AFB, DC

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
6.1102F	2304		

11. TITLE *(Include Security Classification)* The Independence Assumption for a Series or Parallel System when Component Lifetimes are Exponential

12. PERSONAL AUTHOR(S)
John P. Klein and M. L. Moeschberger

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Final	FROM 9-1-82 TO 12-31-87	May 31, 1988	6

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB. GR.
XXXXXXXXXXXXXXX		

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
Exponential component life, Marshall-Olkin bivariate exponential, Freund bivariate exponential, Modeling series (Continued)

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

A common assumption made in modeling system life from series and parallel systems is that the component lives are independent. This study investigates the magnitude of the errors one may incur by erroneously assuming the component lifetimes have independent exponential distributions when in fact the lifetimes follow the bivariate exponential distribution of Marshall & Olkin (series or parallel systems) or that of Freund (parallel systems).

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
Brian W. Woodruff, Major	(202) 767-5027	AFOSR/NM

Block 18- Subject terms (Continued)

systems, Modeling parallel systems, Mean system life.

APPENDIX D

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY NA	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public Release; Distribution Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE NA	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION The Ohio State University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NM
6c. ADDRESS (City, State and ZIP Code) 1314 Kinnear Road Columbus, Ohio 43212	7b. ADDRESS (City, State and ZIP Code) Bldg. 410 Bolling AFB, DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-82-0307
8c. ADDRESS (City, State and ZIP Code) Bldg. 410 Bolling AFB, DC	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO.: 6.1102F; PROJECT NO.: 2304; TASK NO.: ; WORK UNIT NO.:

11. TITLE (Include Security Classification)
Independent or Dependent Competing Risks: Does It Make A Difference?

12. PERSONAL AUTHOR(S)
John P. Klein and M. L. Moeschberger

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Final	FROM 9-1-82 TO 12-31-87	May 31, 1988	36

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	Competing risks; Component life; Modeling series system; Robustness studies; System reliability; Gumbel bivariate exponential; Downton bivariate exponential (Continued)
XXXXXXXXXXXXXX			

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This article investigates the consequences of departures from independence when the component lifetimes in a series system are exponentially distributed. Such departures are studied when the joint distribution is assumed to follow either one of the three Gumbel bivariate exponential models, the Downton bivariate exponential model, or the Oakes bivariate exponential model. Two distinct situations are considered. First, in theoretical modeling of series systems, when the distribution of the component lifetimes is assumed, one wishes to compute system reliability and mean system life. Second, errors in parametric and nonparametric estimation of component reliability and component mean life are studied based on life-test data collected on series systems when the assumption of independence is made erroneously. In both instances, one may be appreciably misled by falsely assuming independent component lifetimes. The amount of error incurred depends upon the correlation between lifetimes and the relative mean life of the two components. In the modeling problem, the level of reliability (Continued)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Brian W. Woodruff, Major	22b. TELEPHONE NUMBER (Include Area Code) (202) 767-5027 22c. OFFICE SYMBOL AFOSR/NM

Block 18- Subject terms (Continued)

Oakes bivariate exponential.

Block 19- Abstract (Continued)

and the length of mean system life also affects the error. In the estimation problem sample size may be influential in determining the magnitude of the error.

The Independence Assumption for a Series or Parallel System when Component Lifetimes are Exponential

John P. Klein
The Ohio State University
Columbus
M. L. Moeschberger
The Ohio State University
Columbus

Key Words—Exponential component life, Marshall - Olkin bivariate exponential, Freund bivariate exponential, Modeling series systems, Modeling parallel systems, Mean system life.

Reader Aids—
Purpose: Widen the state of the art
Special Math needed for explanations: Probability and statistics
Special Math needed to use results: Statistics
Results useful to: Statisticians, reliability engineers, and analysts

Abstract—A common assumption made in modeling system life from series and parallel systems is that the component lives are independent. This study investigates the magnitude of the errors one may incur by erroneously assuming the component lifetimes have independent exponential distributions when in fact the lifetimes follow the bivariate exponential distribution of Marshall & Olkin (series or parallel systems) or that of Freund (parallel systems).

1. INTRODUCTION

First, we shall consider a 1-out-of-2:F system. Such a system functions if and only if both components function. Suppose that if the two components are tested separately, the respective times to failure are X_1 and X_2. A common assumption made in life-testing is that these component failure times follow exponential distributions [2, 3, 4, 5, 7]. Furthermore the assumption that X_1, X_2 are *s*-independent (the prefix *s*- implies "statistically") is often-times invoked. In general, such an assumption is not testable due to the identifiability dilemma [1], ie, if we observe $T \equiv$ minimum $\{X_1, X_2\}$ and $I \equiv \chi\{X_1 < X_2\}$, where $\chi(\cdot)$ denotes the indicator function, then Tsiatis (9) and others have shown that the pair (T, I) provides insufficient information to determine the joint distribution of X_1, X_2. That is, there exists both an *s*-independent and *s*-dependent model for (X_1, X_2) which produces the same joint distribution for (T, I). However, these equivalent *s*-independent and *s*-dependent joint distributions may have quite different marginal distributions. Also due to this identifiability problem there may be several *s*-dependent models with different marginal structures which yield the same observable information, (T, I).

The first purpose of this study, as reported in section 3, is to investigate the effects of departures from this *s*-independence assumption on modeling system life in such a series system when the specific form of departure from *s*-independence is that the joint distribution of (X_1, X_2) is the bivariate exponential distribution of Marshall and Olkin (8). This model postulates the possibility of simultaneous failure of the two components due to a shock simultaneously felt by both components or because one component (say, a rocket booster) explodes and the other component (say, a space shuttle) is destroyed by the explosion.

Secondly, we consider a 2-out-of-2:F system. Such a system functions as long as one of the components functions. We investigate in section 4 the effects of similar departures from the *s*-independence assumption on modeling system life when the specific departures from *s*-independence follow either the Freund (6) bivariate exponential or that of Marshall & Olkin. The Freund model introduces *s*-dependence between X_1 and X_2 such that the failure of the first component changes the parameter of the exponential life distribution of the second component from λ_2 to θ_2 and similarly if the second component would fail first.

2. PROBLEM STATEMENT

Notation

X_i	lifetime of component i
$\bar{F}_I(t)$	system Sf under *s*-independence
$\bar{F}_D(t)$	system Sf under *s*-dependence
μ_I	mean system life under *s*-independence
μ_D	mean system life under *s*-dependence
$t_{p,I}$	point at which system reliability is p under *s*-independence, ie, $p = \bar{F}_I(t_{p,I})$
$t_{p,D}$	point at which system reliability is p under *s*-dependence, ie, $p = \bar{F}_D(t_{p,D})$.
BVE	bivariate exponentiality

Other, standard notation is given in "Information for Readers & Authors" at rear of each issue.

2.1 Model Assumptions

1. Based on testing each component separately, an investigator knows the marginal Sf (t) of the lifetime of i component to be $\exp(-\lambda_i t)$.

2. When the components are installed in a series (1-out-of-2:F) or parallel (2-out-of-2:F) system, the joint Sf$\{x_1, x_2\}$ of component lifetimes may follow the Marshall-Olkin (8) model:

$$\Pr\{X_1 > x_1, X_2 > x_2\} = \exp(-\lambda_1 x_1 - \lambda_2 x_2 - \lambda_{12} \cdot \max\{x_1, x_2\}), \text{ for } \lambda_1, \lambda_2 > 0, \lambda_{12} \geq 0, x_1, x_2 > 0.$$

3. When the components are installed in a 2-out-of-2:F system the joint pdf(X_1, X_2) of component lifetimes may follow the Freund (6) model

$$f(x_1, x_2) = \begin{cases} \lambda_1 \theta_2 \exp\{-\theta_2 x_2 - (\lambda_1 + \lambda_2 - \theta_2)x_1\}, & \text{if } 0 < x_1 < x_2 \\ \lambda_2 \theta_1 \exp\{-\theta_1 x_1 - (\lambda_1 + \lambda_2 - \theta_1)x_2\}, & \text{if } 0 < x_2 < x_1, \end{cases}$$

for $\lambda_1, \lambda_2, \theta_1, \theta_2 > 0$.

4. Under the assumption of s-independence in a 1-out-of-2:F system the system reliability at a mission time t is:

$$\Pr\{\min(X_1, X_2) > t \mid \text{"Independence"}\} = \overline{F}_I(t) = \exp[-(\lambda_1 + \lambda_2)t],$$

mean system life $= \mu_I = 1/(\lambda_1 + \lambda_2)$

$t_{p,I} = -\ln p/(\lambda_1 + \lambda_2)$.

5. Under the assumption of Marshall-Olkin BVE in a 1-out-of-2:F system the system reliability at a mission time t is:

$$\Pr\{\min(X_1, X_2) > t \mid \text{"BVE"}\} = \overline{F}_D(t) = \exp(-\lambda t)$$

$$\lambda \equiv \lambda_1 + \lambda_2 + \lambda_{12},$$

mean system life $= \mu_D = 1/\lambda$,

$t_{p,D} = -\ln p/\lambda$.

6. Under the assumption of s-independence in a 2-out-of-2:F system the system reliability at a mission time t is:

$$\Pr\{\max(X_1, X_2) > t \mid \text{"Independence"}\} = \overline{F}_I(t) = \exp(-\lambda_1 t) + \exp(-\lambda_2 t) - \exp(-(\lambda_1 + \lambda_2)t),$$

mean system life is $\mu_I = 1/\lambda_1 + 1/\lambda_2 - 1/(\lambda_1 + \lambda_2)$

$$= \frac{(\lambda_1 + \lambda_2)^2 - \lambda_1\lambda_2}{\lambda_1\lambda_2(\lambda_1 + \lambda_2)}$$

and t_p must be solved by numerical techniques.

7. Under the assumption of Marshall-Olkin BVE in a 2-out-of-2:F system the system reliability is:

$$\Pr\{\max(X_1, X_2) > t \mid \text{"BVE"}\} = \overline{F}_D(t) = \exp(-(\lambda_1 + \lambda_{12})t) + \exp(-(\lambda_2 + \lambda_{12})t) - \exp(-\lambda t),$$

mean system life is:

$$\mu_D = \frac{(\lambda + \lambda_{12})}{(\lambda_1 + \lambda_{12})(\lambda_2 + \lambda_{12})} - \frac{1}{\lambda},$$

and t_p must be solved by numerical techniques.

8. Under the assumption of Freund BVE in a 2-out-of-2:F system the system reliability is:

$$\overline{F}_D(t) = \frac{\lambda_1 \exp(-\theta_2 t)}{(\lambda_1 + \lambda_2 - \theta_2)} + \frac{\lambda_2 \exp(-\theta_1 t)}{(\lambda_1 + \lambda_2 - \theta_1)} - \frac{(\lambda_1\theta_2 - \theta_1\theta_2 + \theta_1\lambda_2)\exp(-(\lambda_1 + \lambda_2)t)}{(\lambda_1 + \lambda_2 - \theta_2)(\lambda_1 + \lambda_2 - \theta_1)}$$

and mean system life is:

$$\mu_D = \lambda_1/[\theta_2(\lambda_1 + \lambda_2 - \theta_2)] + \lambda_2/[\theta_1(\lambda_1 + \lambda_2 - \theta_1)] - (\lambda_1\theta_2 - \theta_1\theta_2 + \theta_1\lambda_2)/[(\lambda_1 + \lambda_2)(\lambda_1 + \lambda_2 - \theta_2) \cdot (\lambda_1 + \lambda_2 - \theta_1)], \text{ if } \lambda_1 + \lambda_2 \neq \theta_2 \text{ and } \lambda_1 + \lambda_2 \neq \theta_1$$

For the special instances when $\lambda_1 + \lambda_2 = \theta_2$ and/ or $\lambda_1 + \lambda_2 = \theta_1$, the pdf in model assumption 3 is slightly simplified and the following different equations for system reliability and mean system life must be used.

If $\lambda_1 + \lambda_2 \neq \theta_2$, and $\lambda_1 + \lambda_2 = \theta_1$:

$$\overline{F}_D(t) = \lambda_2 t \exp(-(\lambda_1 + \lambda_2)t) + \frac{\exp(-\theta_2 t)}{(\lambda_1 + \lambda_2 - \theta_2)} + \frac{\{(\lambda_1 + \lambda_2)(\lambda_2 - \lambda_1\theta_2) - \lambda_2\theta_2\}\exp(-(\lambda_1 + \lambda_2)t)}{(\lambda_1 + \lambda_2)(\lambda_1 + \lambda_2 - \theta_2)},$$

$$\mu_D = \lambda_2/(\lambda_1 + \lambda_2)^2 + 1/[\theta_2(\lambda_1 + \lambda_2 - \theta_2)] + \frac{(\lambda_1 + \lambda_2)(\lambda_2 - \lambda_1\theta_2) - \lambda_2\theta_2}{(\lambda_1 + \lambda_2)^2(\lambda_1 + \lambda_2 - \theta_2)}.$$

If $\lambda_1 + \lambda_2 = \theta_2$, and $\lambda_1 + \lambda_2 \neq \theta_1$:

$$\overline{F}_D(t) = \lambda_1 t \exp(-(\lambda_1 + \lambda_2)t) + \frac{\exp(-\theta_1 t)}{(\lambda_1 + \lambda_2 - \theta_1)} + \frac{\{(\lambda_1 + \lambda_2)(\lambda_1 - \lambda_2\theta_1) - \lambda_1\theta_1\}\exp(-(\lambda_1 + \lambda_2)t)}{(\lambda_1 + \lambda_2)(\lambda_1 + \lambda_2 - \theta_1)},$$

$$\mu_D = \lambda_1/(\lambda_1 + \lambda_2)^2 + 1/[\theta_1(\lambda_1 + \lambda_2 - \theta_1)] + \frac{(\lambda_1 + \lambda_2)(\lambda_1 - \lambda_2\theta_1) - \lambda_1\theta_1}{(\lambda_1 + \lambda_2)^2(\lambda_1 + \lambda_2 - \theta_1)}.$$

If $\lambda_1 + \lambda_2 = \theta_1 = \theta_2$:

$$\overline{F}_D(t) = \{1 + (\lambda_1 + \lambda_2)t\}\exp(-(\lambda_1 + \lambda_2)t),$$

$$\mu_D = 2/(\lambda_1 + \lambda_2).$$

As before, t_p must be evaluated by numerical techniques.

The problem to be addressed is: How far off may a researcher be in modeling system reliability, mean system life, and mission time at which system reliability is p for series systems (Marshall-Olkin BVE) and for parallel systems (Marshall-Olkin BVE or Freund BVE).

3. MODELING EFFECTS OF DEPARTURES FROM s-INDEPENDENCE IN SERIES SYSTEMS

From model assumptions 4 and 5, the relative error in modeling system reliability under Marshall-Olkin BVE is

$$\frac{\bar{F}_D(t) - \bar{F}_I(t)}{\bar{F}_I(t)} = \frac{\exp(-\lambda t) - \exp(-(\lambda_1 + \lambda_2)t)}{\exp(-(\lambda_1 + \lambda_2)t)}$$

$$= \exp(-\lambda_{12}t) - 1$$

$$= \exp\left\{\frac{-\rho\,(\lambda_1 + \lambda_2)t)}{(1 - \rho)}\right\} - 1$$

$\rho = \lambda_{12}/\lambda$ is the correlation between X_1 and X_2.

This shows explicitly that the relative error ranges from 0 (when $\rho = 0$) to $-\infty$ (when $\rho = +1$). Thus assuming s-independence consistently overestimates system reliability and in some instances the results may be very far off. When mission time is $t_p = -\ln p/(\lambda_1 + \lambda_2)$ the relative error is $\exp\left(\frac{\rho}{1-\rho}\ln p\right) - 1 = p^{\rho/(1-\rho)} - 1$. From model assumptions 4 and 5, the error in modeling mean system life is:

$$\frac{\mu_D - \mu_I}{\mu_I} = \frac{1/\lambda - 1/(\lambda_1 + \lambda_2)}{1/(\lambda_1 + \lambda_2)} = -\frac{\lambda_{12}}{\lambda} = -\rho.$$

Again s-independence consistently overestimates the mean system life.

4. MODELING EFFECTS OF DEPARTURES FROM INDEPENDENCE IN PARALLEL SYSTEMS

4.1 Under Marshall-Olkin BVE.

From model assumptions 6 and 7, the relative error is:

$$\frac{\bar{F}_D(t) - \bar{F}_I(t)}{\bar{F}_I(t)} = \exp(-\lambda_{12}t) - 1$$

$$= \exp\{-\frac{\rho}{1-\rho}(\lambda_1 + \lambda_2)t\} - 1$$

which interestingly enough is the same relative error obtained in section 3 and the same comments apply here.

The relative error in modeling mean system life is:

$$\frac{\mu_D - \mu_I}{\mu_I} = -\rho\,\frac{(k^4+2k^3+k^2+2k+1)+\rho\,(k^3+4k^2+k)-k^2\rho^2}{(k^2+k+1)(1+k\rho)(k+\rho)}$$

$k = \lambda_1/\lambda_2$.

Again s-independence consistently overestimates mean system life; the magnitude of this error is plotted in figure 1.

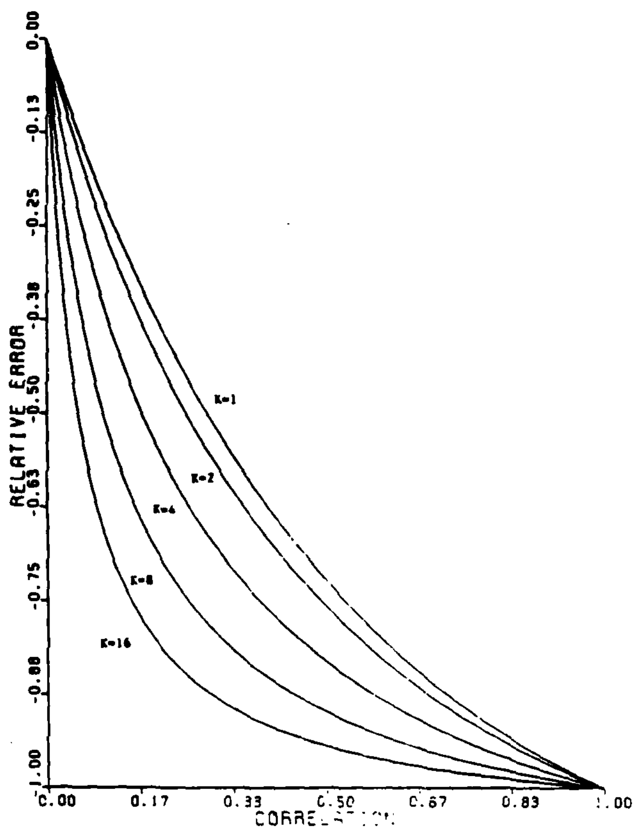


Fig. 1. Relative Error in Modeling Mean System Life Under Marshall-Olkin BVE for $\lambda_1 = k\,\lambda_2$

4.2 Under Freund BVE.

From model assumptions 6 and 8 the relative errors of mean system life and system reliability are:

$$\frac{\mu_D - \mu_I}{\mu_I} \text{ and } \frac{\bar{F}_D(t) - \bar{F}_I(t)}{\bar{F}_I(t)};$$

though not simple expressions, they may be easily calculated. Figure 2 shows plots of the relative error of mean system life for some representive values of λ_i and θ_i, viz, $\lambda_1 = 1$; $\lambda_2 = 0.5, 1.0, 2.0$; $0.25 \leqslant \theta_1 \leqslant 4.0$ (on horizontal axis); and $\theta_2 = 0.25, 0.5, 1.0, 2.0, 4.0$. Figure 3 shows plots

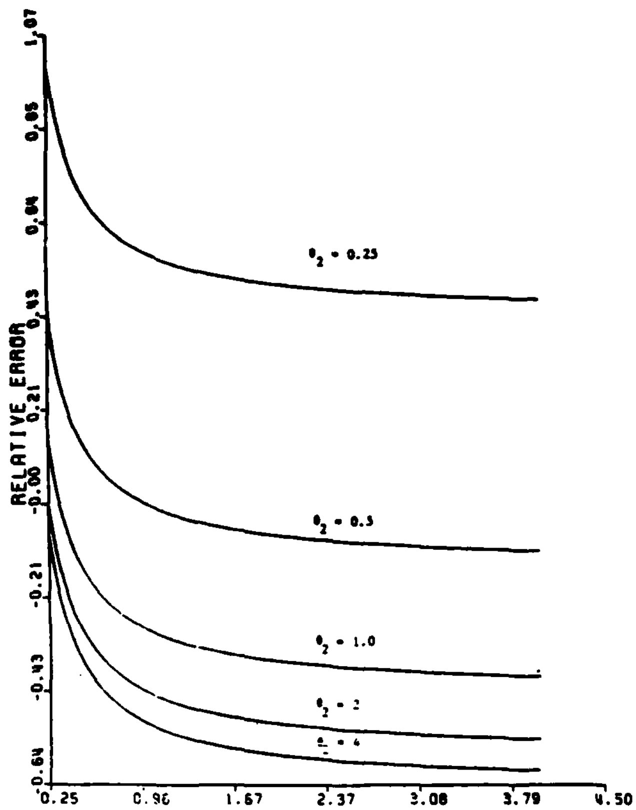


Fig. 2A. Relative Error of Mean System Life Under Freund BVE for $\lambda_1 = 1$; $\lambda_2 = .5$.

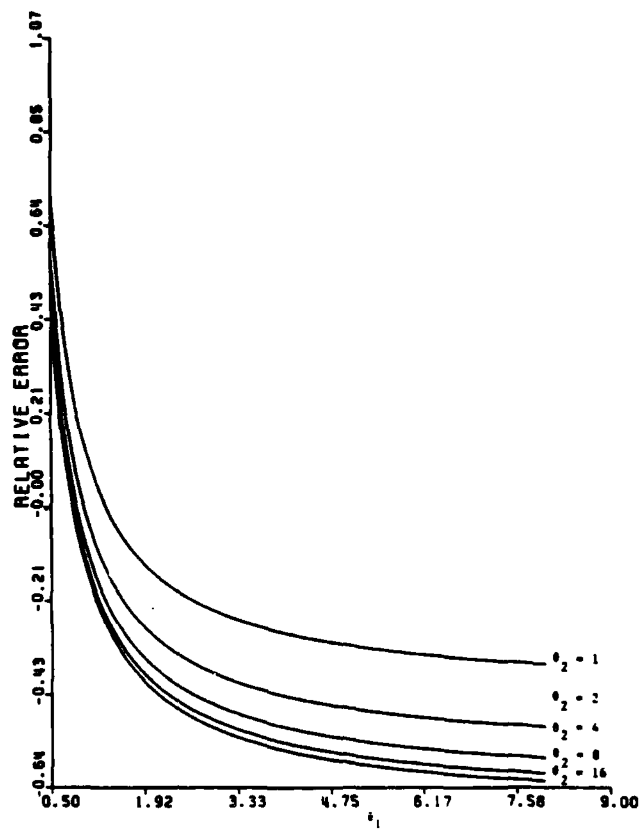


Fig. 2C. Relative Error of Mean System Life Under Freund BVE for $\lambda_1 = 1$; $\lambda_2 = 2$.

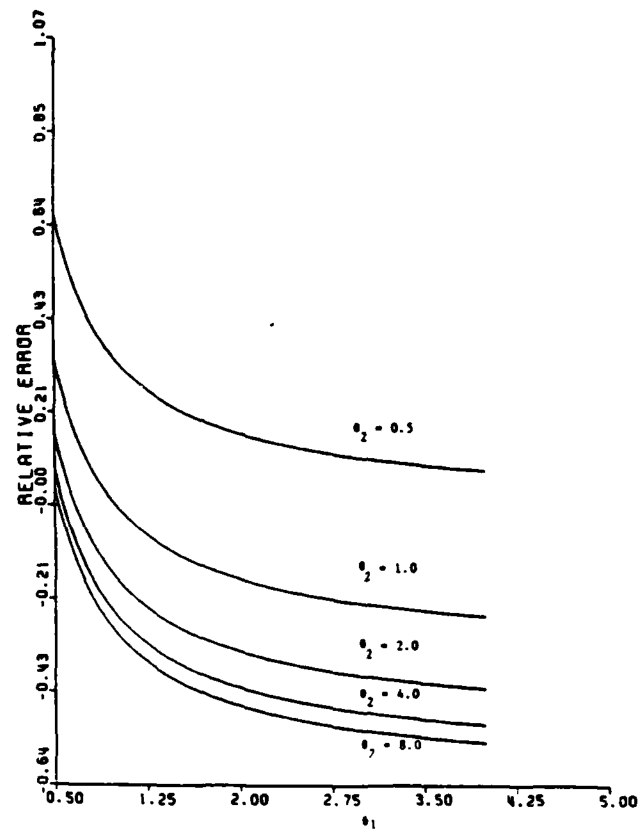


Fig. 2B. Relative Error of Mean System Life Under Freund BVE for $\lambda_1 = 1$; $\lambda_2 = 1$.

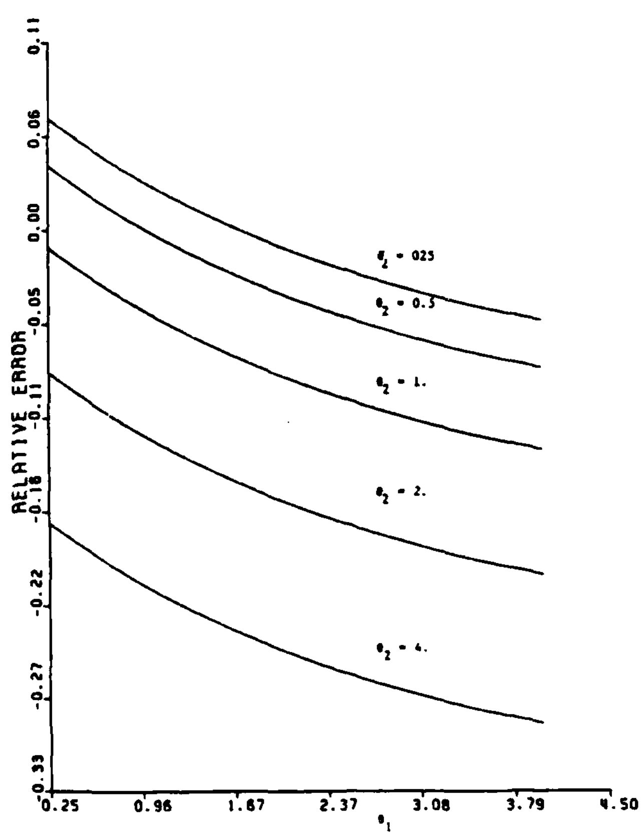


Fig. 3A. Relative Error of System Reliability Under Freund BVE for $t_{.90} = .070$, $\lambda_1 = 1$, $\lambda_2 = .5$.

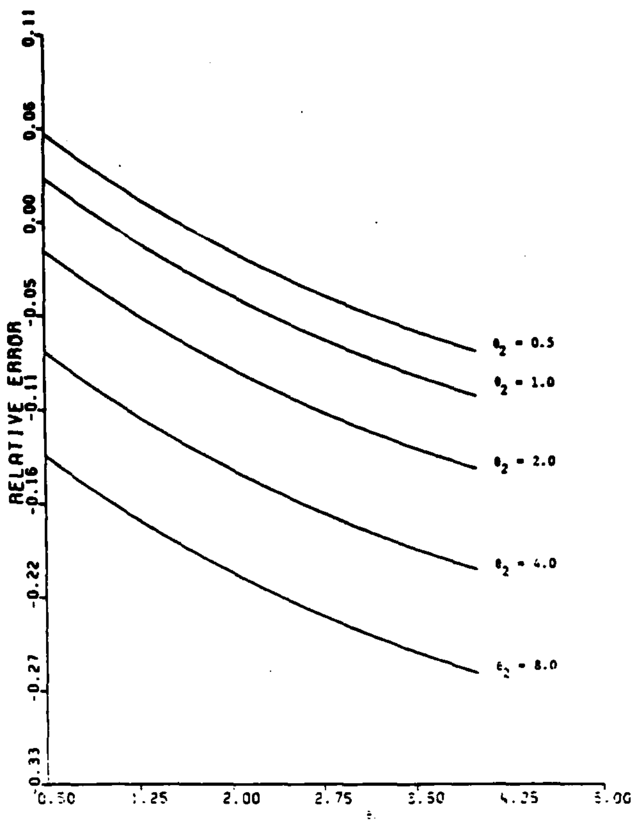


Fig. 3B. Relative Error of System Reliability Under Freund BVE for $t_{.90} = .053$, $\lambda_1 = 1$, $\lambda_2 = 1$.

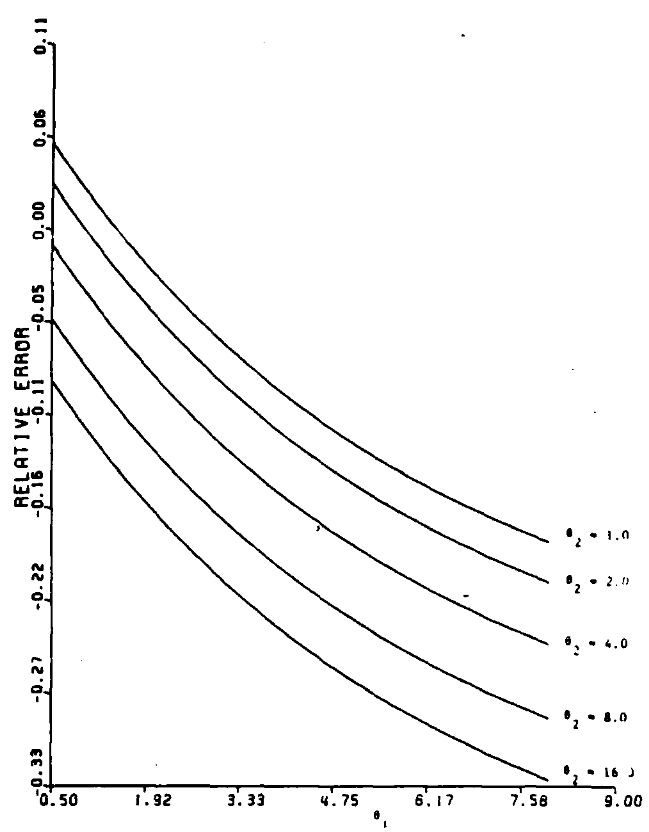


Fig. 3C. Relative Error of System Reliability Under Freund BVE for $t_{.90} = .035$, $\lambda_1 = 1$, $\lambda_2 = 2$.

of the relative error of system reliability for $t_{.9} = -\ln(.90)/(\lambda_1 + \lambda_2)$ for the same values of λ_i and θ_i. For the case of *s*-independence, $\lambda_1 = \theta_1 = 1$ and $\lambda_2 = \theta_2 = 0.5, 1, 2$, the relative error is, of course, zero in the plots. For situations in which failure of the first component causes the hazard rate θ_2 of the second component to be increased (or decreased), *s*-independence overestimates (or underestimates) mean system life and system reliability. This overestimate (or underestimate) becomes substantial if the hazard rate θ_2 is dramatically increased (or decreased) over the hazard rate λ_2 of the second component tested separately. A similar comment can be made for θ_1 and λ_1. Positive (or negative correlation) of the component lifetimes is a consequence of the hazard rate θ_i being greater (or less) than the hazard rate λ_i of component *i* tested separately. The magnitude of the relative error of mean system life and system reliability may easily be found in the figures 2 & 3, respectively.

ACKNOWLEDGMENT

This research was supported under contract AFOSR-82-0307 for the U.S. Air Force Office of Scientific Research.

REFERENCES

[1] A.P. Basu, J.P. Klein, "Some recent results in competing risks", *Survival Analysis*, Crowley and Johnson eds., IMS, 1982, 210-229.

[2] T.J. Boardman, P.J. Kendell, "Estimation in compound exponential failure models", *Technometrics*, vol. 12, 1970, pp. 891-900.

[3] A. Chao, "Approximate mean squared errors of estimators of reliability for k-out-of-m systems in the independent exponential case", *J. Amer. Statist. Assoc.*, vol. 76, 1981, pp. 720-724.

[4] H.A. David, M.L. Moeschberger, *The Theory of Competing Risks*, Macmillan Publishing Co., 1978.

[5] B. Epstein, M. Sobel, "Some theorems relevant to life testing from an exponential distribution", *Ann. Math Statist*, vol. 25, 1954, pp. 373-381.

[6] J.E. Freund, "A bivariate extension of the exponential distribution", *J. Amer. Statist. Assoc.*, vol. 56, 1961, pp. 971-977.

[7] N.R. Mann, "Simplified expressions for obtaining approximately optimum system reliability confidence bounds from exponential subsystem data", *J. Amer. Stat. Assoc.*, vol. 69, 1974, pp., 492-495.

[8] A.W. Marshall, I. Olkin, "A multivariate exponential distribution", *J. Amer. Statist. Assoc.*, vol. 62, 1967, pp. 30-40.

[9] A.A. Tsiatis, "A nonidentifiability aspect of the problem of competing risks", *Proc. Nat. Acad. Sci.*, vol. 72, 1977, pp. 20-22.

AUTHORS

Dr. John P. Klein; Department of Statistics; Ohio State University; Columbus, Ohio 43210 USA.

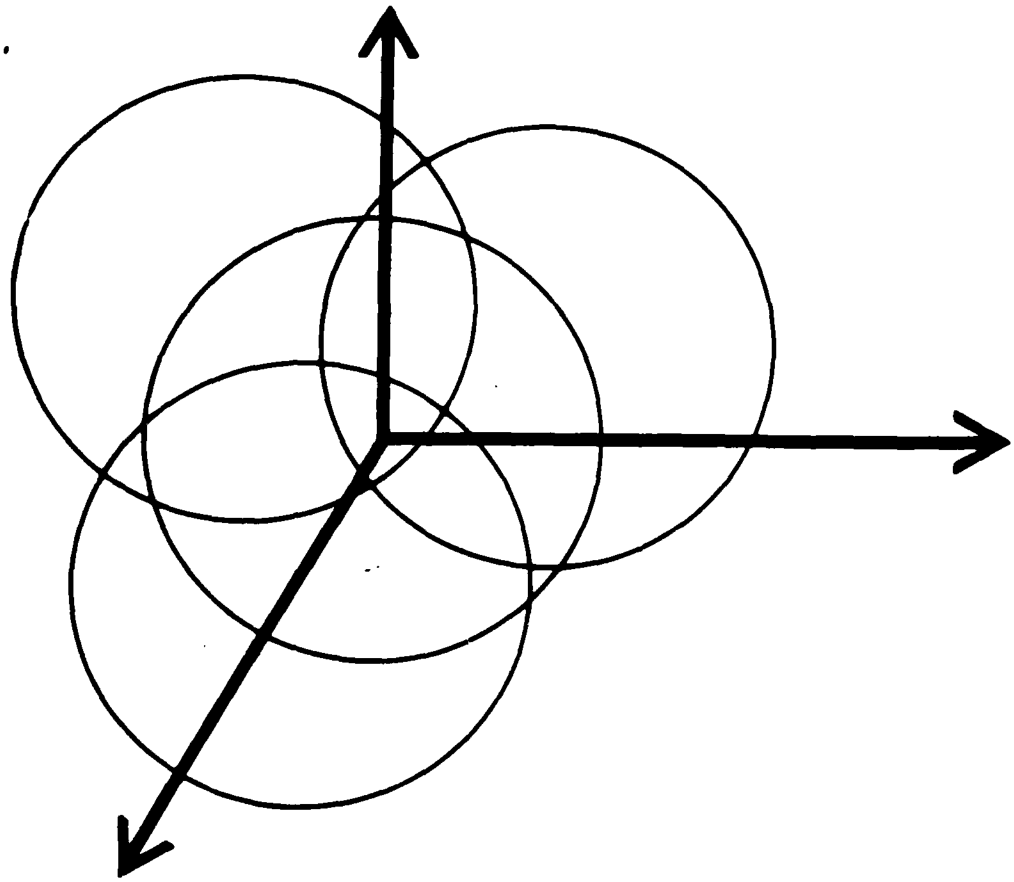
John P. Klein was born in Milwaukee, Wisconsin in 1950. He received a BA and MS in mathematics at the University of Wisconsin at Milwaukee and a PhD in statistics from the University of Missouri at Columbia. He is an Assistant Professor in the Department of Statistics at The Ohio State University and the Biostatistician for The Ohio State University Comprehensive Cancer Center.

Dr. Melvin L. Moeschberger; Department of Preventive Medicine; Ohio State University; Columbus, Ohio 43210 USA.

Melvin L. Moeschberger was born in Berne, Indiana in 1940. He received a BS in mathematics and chemistry from Taylor University, an MS in mathematics from Ohio University, and a PhD in statistics from North Carolina State University. He is an Associate Professor in the Department of Preventive Medicine and is Director of the Biometrics Laboratory in the College of Medicine at The Ohio State University.

Manuscript TR83-113 received 1983 July 30; revised 1986 March 28.

★ ★ ★



communications in statistics

Communications in Statistics
Part A: Theory and Methods
Part B: Simulation and Computation
Part C: Stochastic Models

Communications in Statistics is a multi-part journal.

Part A: Theory and Methods focuses primarily on papers describing new applications of known statistical methods to actual problems in industry and government and on articles with a strong mathematical orientation that are significant to statistical studies. In addition, ***Part A*** also offers communications that discuss practical problems with either only ad hoc solutions or none at all; in cases where there is a difference of opinion on particular techniques, all parties involved vigorously debate the issue with thought provoking commentaries.

Part B: Simulation and Computation deals specifically with problems at the interface of statistics and computer science, including tables of, and algorithms for, statistical functions and numerical solutions to outstanding problems, whether by simulation or the use of special functions. Papers are generally application oriented, although when practical utility is demonstrated, the journal presents theoretical papers on appropriate topics.

Part C: Stochastic Models (Affiliated Publication of the Operations Research Society of America) (ORSA) offers an interdisciplinary presentation on the uses of probability theory with contributions on mathematical methodology ranging from structural, analytic, and algorithmic to experimental approaches. This publication discusses the practical applications of stochastic models to such diverse areas as biology, computer science,telecommunication, modeling, inventories and dams, reliability, storage, quewing theory, and operations research.

All three journals stress practicality and innovation, and their direct reproduction format ensures truly rapid communication of the very newest ideas, problems and solutions in all areas of the field, keeping readers at the forefront of statistical inquiry.

COMMUN. STATIST. -SIMULA., 16(2), 507-533 (1987)

INDEPENDENT OR DEPENDENT COMPETING RISKS: DOES IT MAKE A DIFFERENCE?

John P. Klein
Department of Statistics
The Ohio State University
Columbus, Ohio 43210

M.L. Moeschberger
Department of Preventive Medicine
The Ohio State University
Columbus, Ohio 43210

Key Words and Phrases: competing risks; component life; modeling series systems; robustness studies; system reliability; Gumbel bivariate exponential; Downton bivariate exponential; Oakes bivariate exponential.

ABSTRACT

This article investigates the consequences of departures from independence when the component lifetimes in a series system are exponentially distributed. Such departures are studied when the joint distribution is assumed to follow either one of the three Gumbel bivariate exponential models, the Downton bivariate exponential model, or the Oakes bivariate exponential model. Two distinct situations are considered. First, in theoretical modeling of series systems, when the distribution of the component lifetimes is assumed, one wishes to compute system reliability and mean system life. Second, errors in parametric and nonparametric estimation of component reliability and component mean life are studied based on life-test data collected on series systems when the assumption of independence is made

erroneously. In both instances, one may be appreciably misled by falsely assuming independent component lifetimes. The amount of error incurred depends upon the correlation between lifetimes and the relative mean life of the two components. In the modeling problem, the level of reliability and the length of mean system life also affects the error. In the estimation problem sample size may be influential in determining the magnitude of the error.

INTRODUCTION

Consider a system consisting of two components linked in series. For such a system the failure of either of the components causes the system to fail. In a biological or medical context the components can be different lethal diseases and/or different reasons for removal from a study. In a clinical trials framework the primary response of interest, death or remission, and censoring can be considered as components of the system. This general formulation has been detailed in the theory of competing risks (cf. David and Moeschberger (1978)).

A common assumption in such a formulation is that the component lifetimes are statistically independent. Several authors have shown that based on data from series systems only, this assumption, by itself, is not testable because there is no way to distinguish between independent or dependent component lifetimes (see Basu (1981), Basu and Klein (1982), Miller (1977), Peterson (1976), etc.). However, several authors (see Lagakos (1979) p. 152 and Easterling (1980) p. 131) have pointed out the need to determine, quantitatively, how far off one might be if an analysis is based on an incorrect assumption of independence.

To study the effects of erroneously assuming independence we shall assume that each of two component lifetimes is exponentially distributed when tested separately and that the property of marginal exponentiality will be preserved even though some dependence may be induced when the components are linked in series. The assumption of exponentially distributed component lifetimes has been made by Mann and Grubbs (1974) when finding confidence bounds on system reliability, Boardman and Kendell (1970) when estimating component lifetimes from system data, and Miyamura (1982) when combining component and system data. (See Barlow and Proschan (1975) or Mann, Schafer, and Singpurwalla (1974) for a more complete review.) We shall model the dependence structure by the three models of Gumbel (1960), a model proposed by Downton (1970), and a model described by Oakes (1982). These models are briefly described in Section 1. The effects of a departure from the assumption of independent component lifetimes will be addressed for two distinct situations.

The first situation arises in modeling the performance of a theoretical series system constructed from two components. Here, based on testing each component separately or on engineering design principles, it is reasonable to assume that the component lifetimes are exponentially distributed with known parameter values. Based on this information, we wish to predict parameters such as the mean life or reliability of a series system constructed from these components. In Section 2 we describe how these quantities are affected by departures from independence.

The second situation involves making inferences about component lifetime distributions from data collected on series systems. Commonly, data collected on such systems are analyzed by assuming a constant-sum model, of which independence is a special case (compare Williams and Lagakos (1977) and Lagakos and Williams (1978)). In Section 3.1 we study the properties of the maximum likelihood estimators of the component mean life calculated under an erroneous assumption of independent exponential component lifetimes as mentioned above. Because of the widespread use of the nonparametric estimator for component reliability proposed by Kaplan-Meier (1958) we study, in Section 3.2, the properties of this estimator when the marginal reliabilities are exponential and independence is incorrectly assumed.

1. The Models

Consider a two component series system with component life lengths X_1, X_2. Suppose that each X_i has an exponential survival function

$$\bar{F}_i(x) = P(X_i > x) = \exp(-\lambda_i x), \quad \lambda_i \geq 0,\ x \geq 0. \tag{1.1}$$

This assumption is made on the basis of extensive testing of each component separately or on knowledge of the underlying mechanism of failure.

To examine the effects of a departure from independence we consider five bivariate exponential models, each with marginals equivalent to (1.1). The first three models are due to Gumbel (1960); the last two models are due to Downton (1970) and Oakes (1982).

1.1 Gumbel's Model A

For this model the joint survival function is

$$P(X_1 > x_1, X_2 > x_2) = \exp(-\lambda_1 x_1 - \lambda_2 x_2 - \lambda_{12} x_1 x_2),$$
$$x_1, x_2 \geq 0, \quad \lambda_1, \lambda_2 > 0, \ 0 \leq \lambda_{12} \leq \lambda_1 \lambda_2. \tag{1.2}$$

The correlation between X_1, X_2 is

$$\rho = -\frac{\lambda_1 \lambda_2}{\lambda_{12}} \exp(\lambda_1 \lambda_2 / \lambda_{12}) E_i(-\lambda_1 \lambda_2 / \lambda_{12}) - 1,$$

where $E_i(z) = \int_{-z}^{\infty} \frac{\exp(-u)}{u}\, du$ is the integrated logarithm.

For this model ρ varies from $-.40365$ to 0 as λ_{12} decreases from $\lambda_1 \lambda_2$ to 0. It is never positive. The regression X_1 on X_2 is nonlinear with

$$E(X_1 | X_2 = x_2) = (\lambda_1 \lambda_2 + \lambda_2 \lambda_{12} x_2 - \lambda_{12}) / (\lambda_1 + \lambda_{12} x_2)^2.$$

1.2 Gumbel's Model B

For this model the joint survival function is

$$P(X_1 > x_1, X_2 > x_2) = \exp(-\lambda_1 x_1 - \lambda_2 x_2)\{1 + 4\rho\,(1-\exp(-\lambda_1 x_1))\,(1-\exp(-\lambda_2 x_2))\},$$
$$\lambda_1,\ \lambda_2 > 0,\ x_1,\ x_2 \geq 0,\ -1/4 \leq \rho \leq 1/4. \qquad (1.3)$$

The correlation, ρ , may be positive or negative. The regression of X_1 on X_2 is again nonlinear with

$$E(X_1 | X_2 = x_2) = \{1 + 2\rho - 4\rho \exp(-\lambda_2 x_2)\}/\lambda_1.$$

The effects of a departure from independence on modeling system reliability and estimating component reliabilities has been studied in detail in Moeschberger and Klein (1984) for this model.

1.3 Gumbel's Model C

For this model the joint survival function is

$$P(X_1 > x_1, X_2 > x_2) = \exp\{-((\lambda_1 x_1)^m + (\lambda_2 x_2)^m)^{1/m}\},$$
$$\lambda_1,\ \lambda_2 > 0,\ m \geq 1,\ x_1,\ x_2 \geq 0. \qquad (1.4)$$

The correlation is

$$\rho = (4 + 2m) \int_0^{\Pi/2} \left[\frac{(\cos\theta \sin\theta)^m}{(\cos\theta^m + \sin\theta^m)^{2+2/m}} \right] d\theta - 1$$

which varies from 0 to 1. For this model $m = 1$ corresponds to independence and as $m \to \infty$

$$P(X_1 > x_1, X_2 > x_2) \to \text{minimum}\ (\exp(-\lambda_1 x_1), \exp(-\lambda_2 x_2)), \qquad (1.5)$$

the Fréchet (1958) upper bound for these marginals.

1.4 Downton's Model

Downton (1970) suggests modeling bivariate exponential systems by a successive damage model. This model assumes that in a two component system the times between successive shocks on each component have independent exponential distributions and that the number of shocks required to cause each component to fail follows a bivariate geometric distribution. The joint probability density function of the component lifetimes is

$$f(x_1,x_2) = \frac{\lambda_1\lambda_2}{1-\rho}\exp\left[-\left(\frac{(\lambda_1 x_1 + \lambda_2 x_2)}{1-\rho}\right)\right] I_0\left(\frac{2\sqrt{\rho\lambda_1\lambda_2\, x_1 x_2}}{1-\rho}\right), \tag{1.6}$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind of order zero, and $\lambda_1,\ \lambda_2 > 0,\ x_1,\ x_2 \geq 0,\ 0 \leq \rho \leq 1$.
The correlation between X_1, X_2 is ρ which spans the interval (0,1). As $\rho \to 1$ the joint survival function of X_1, X_2 approaches the upper Fréchet distribution (1.5). For this model $E(X_1|X_2 = x_2) = (1-\rho)/\lambda_1 + \rho\lambda_2 x_2/\lambda_1$.

1.5 Oakes' Model

Oakes (1982) has proposed a model for bivariate survival data. This model was first proposed by Clayton (1978) to model association in bivariate lifetables. Special cases of Oakes' general model have been suggested by Lindley and Singpurwalla (1985) and Hutchinson (1981).

For this model the joint survival probability is

$$P(X_1 > x_1,\ X_2 > x_2) = [\exp(\lambda_1(\theta-1)x_1) + \exp(\lambda_2(\theta-1)x_2)-1]^{-1/(\theta-1)}$$

where $\lambda_1,\ \lambda_2 > 0,\ \theta \geq 1,\ x_1,\ x_2 \geq 0.$ (1.7)

For $\theta = 1$, X_1, X_2 are independent and $P(X_1 > x_1,\ X_2 > x_2) \to$ bound in (1.5) as $\theta \to \infty$. For this model Kendall's coefficient of concordance is $\tau = (\theta-1)/(\theta+1)$ which spans the range 0 to 1. The correlation, ρ, also spans the range 0 to 1 and is found numerically.

This model has the following physical interpretation. Let $r(x_1|X_2 = x_2)$ and $r(x_1|X_2 > x_2)$ be the conditional failure rates of X_1 given $X_2 = x_2$ and $X_2 > x_2$, respectively. Then $r(x_1|X_2 = x_2) = \theta r(x_1|X_2 > x_2)$.

The model can also be derived from a random effects model. This formulation assumes that when the components are tested separately unde ideal conditions the component survival functions are

$$S_i(x) = \exp(-\exp(\lambda_i x(\theta-1)) + 1),\ i = 1,\ 2,$$

and that when the two components are put in a series system in the operating environment there is a random factor W which simultaneously changes each component life distribution to $S_1^W(x)$. If W has a gamma distribution with density function $g(w) \propto w^{\frac{1}{\theta-1}-1} e^{-w}$ then, unconditionally, the joint survival function (1.7) holds.

1.6 Fréchet Bounds

Fréchet (1958) obtained bounds of the joint survival functions which can be obtained for any set of marginal distributions. For exponential marginals these are

$$\text{MAXIMUM}\ (e^{-\lambda_1 x_1} + e^{-\lambda_2 x_2} - 1,\ 0) \le P(X_1 > x_1,\ X_2 > x_2)$$

$$\le \text{MINIMUM}\ (e^{-\lambda_1 x_1},\ e^{-\lambda_2 x_2}) .$$

For this set of marginals the lower Fréchet distribution has correlation - .694 and the upper Fréchet distribution has correlation 1.0. These are the minimal and maximal correlations for exponential marginals.

2. Errors in Modeling System Life

Suppose that based on extensive testing or based on theoretical considerations each of the two components in a series system is known to have an exponential distribution, (1.1) with marginal means $1/\lambda_1$, $1/\lambda_2$, respectively. It is of interest to predict the system reliability $\bar{F}(x) = P(X_1 > x, X_2 > x)$ and the system mean life $\mu = \int_0^\infty \bar{F}(t)\,dt$. If the investigator assumes that the two components are independent then the system reliability is

$$\bar{F}_I(x) = \exp(-(\lambda_1 + \lambda_2)x) \tag{2.1}$$

and system mean life is $\mu_I = 1/(\lambda_1 + \lambda_2)$.

If the components are not independent, but in fact follow one of the models in Section 1, then convenient measures of the effects of incorrectly assuming independence are

$\Delta(x) = (\bar{F}(x) - \bar{F}_I(x))/\bar{F}_I(x)$ and $\delta = (\mu - \mu_I)/\mu_I$,

for predicting system reliability and system mean life, respectively. In these, $\bar{F}(x)$ and μ are computed under the appropriate dependent model. Values of $\bar{F}(x)$ can be computed directly for (1.2), (1.3), (1.4), (1.6) (by numerical integration) or from (1.7). Expressions for μ are given in Appendix 1. All expressions for $\Delta(x)$ and δ depend on the values of λ_1 and λ_2 only through the ratio $\lambda_1/\lambda_2 = K$ and, for $K < 1$, the values are equivalent to those for $K' = \frac{1}{K}$. For the upper Fréchet distribution

$$\Delta(x_p) = p^{-\left(\frac{K}{K+1}\right)} - 1 \text{ where } K \geq 1$$

and x_p is the upper percentile of $\bar{F}_I(x)$, i.e., $\bar{F}_I(x_p) = p$. Also $\delta = 1/K$ for $K > 1$. For the lower Fréchet distribution

$$\Delta(x_p) = \begin{cases} p^{-\left(\frac{1}{K+1}\right)} + p^{-\left(\frac{K}{K+1}\right)} - p^{-1} - 1 & \text{if } p^{\left(\frac{K}{K+1}\right)} + p^{\left(\frac{1}{K}\right)} - 1 > 0 \\ -1 & \text{otherwise} \end{cases}$$

and $\delta = \dfrac{K^2 + K + 1}{K} - \dfrac{(K+1)^2}{K} Y + (K+1)\ln(Y)$

where Y is the solution of the equality $X^K + X = 1$. Table I gives the values of $\Delta(x_p) \times 100\%$ and $\delta \times 100\%$ for $p = .9, .7, .5, .3, .1$ for the upper and lower Fréchet distributions.

From Table I we see that the largest percent error occurs when the λ parameters are equal (K=1). Also for fixed K there is relatively small error (smaller than 5.41%) in estimating system reliability by modeling a dependent system by an independent system when $\bar{F}(x)$ is large (say, $\bar{F}(x) > .9$). For small values of system reliability, one can be appreciably misled. Errors in estimating system mean life appear to be substantial unless one component has considerably longer marginal life than the second one. In that instance, one can see instinctively that the correlation would have a minimal impact.

Figures 1A, 1B and 2A, 2B are plots of $\Delta(x_p)$ for $p = .25, .75$ and for $K = \lambda_1/\lambda_2 = 1, .67$, respectively, for the five models described in Section 1 as a function of correlation. It appears that substantial errors may be made in modeling system reliability with moderate amounts of dependence.

TABLE I

UPPER AND LOWER BOUNDS ON THE PERCENT ERROR IN MODELING SYSTEM LIFE

	$\bar{F}(x)=0.9$		$\bar{F}(x)=0.7$		$\bar{F}(x)=0.5$		$\bar{F}(x)=0.3$		$\bar{F}(x)=0.1$		MEAN LIFE	
K	LOWER BOUND	UPPER BOUND	LOWER BOUND	UPPER BOUND	LOWER BOUND	UPPER BOUND	LOWER BOUND	UPPER BOUND	LOWER BOUND	UPPER BOUND	LOWER BOUND	UPPER BOUND
1	-0.29	5.41	-3.81	19.52	-17.16	41.42	-100.00	82.57	100.00	216.23	-38.83	100.00
2	-0.26	3.57	-3.39	12.62	-15.27	25.99	-100.00	49.38	-100.00	115.44	-37.07	50.00
3	-0.22	2.67	-2.86	9.33	-12.90	18.92	-51.52	35.12	-100.00	77.83	-34.85	33.33
4	-0.19	2.13	-2.44	7.39	-11.02	14.87	-44.11	27.23	-100.00	58.49	-32.83	25.00
5	-0.16	1.77	-2.12	6.12	-9.57	12.25	-38.38	22.22	-100.00	46.78	-31.06	20.00
6	-0.14	1.52	-1.87	5.23	-8.45	10.41	-33.91	18.77	-100.00	38.95	-29.53	16.67
7	-0.13	1.33	-1.67	4.56	-7.55	9.05	-30.33	16.24	-100.00	33.35	-28.18	14.29
8	-0.12	1.18	-1.51	4.04	-6.82	8.01	-27.42	14.31	-100.00	29.15	-26.99	12.50
9	-0.11	1.06	-1.37	3.63	-6.22	7.18	-25.02	12.79	-100.00	25.89	-25.93	11.11
10	-0.10	0.96	-1.26	3.30	-5.71	6.50	-22.99	11.57	-100.00	23.28	-24.98	10.00
11	-0.09	0.88	-1.17	3.02	-5.28	5.95	-21.27	10.55	-100.00	21.15	-24.12	9.09
12	-0.08	0.81	-1.08	2.78	-4.91	5.48	-19.78	9.70	-100.00	19.38	-23.34	8.33
13	-0.08	0.76	-1.01	2.58	-4.59	5.08	-18.49	8.98	-100.00	17.88	-22.62	7.69
14	-0.07	0.70	-0.95	2.41	-4.30	4.73	-17.35	8.36	-100.00	16.59	-21.96	7.14
15	-0.07	0.66	-0.90	2.25	-4.05	4.43	-16.35	7.82	-100.00	15.48	-21.35	6.67
16	-0.06	0.62	-0.85	2.12	-3.83	4.16	-15.45	7.34	-100.00	14.50	-20.78	6.25
17	-0.06	0.59	-0.80	2.00	-3.63	3.93	-14.65	6.92	-100.00	13.65	-20.26	5.88
18	-0.06	0.56	-0.76	1.89	-3.45	3.72	13.93	6.54	-100.00	12.88	-19.76	5.56
19	-0.06	0.53	-0.73	1.80	-3.29	3.53	-13.27	6.20	-100.00	12.20	-19.30	5.26
20	-0.05	0.50	-0.69	1.71	-3.14	3.36	-12.67	5.90	-92.26	11.59	-18.87	5.00
21	-0.05	0.48	-0.66	1.63	-3.00	3.20	-12.13	5.63	-88.34	11.03	-18.46	4.76
22	-0.05	0.46	-0.64	1.56	-2.88	3.06	-11.63	5.37	-84.73	10.53	-18.08	4.55
23	-0.05	0.44	-0.61	1.50	-2.76	2.93	-11.16	5.14	-81.41	10.07	17.71	4.35
24	-0.04	0.42	-0.59	1.44	-2.66	2.81	-10.74	4.93	-78.34	9.65	-17.37	4.17
25	-0.04	0.41	-0.57	1.38	-2.56	2.70	-10.34	4.74	-75.49	9.26	-17.04	4.00

Figures 3A, 3B are plots of δ for all five models and for $K = \lambda_1 / \lambda_2 = 1, .67$, respectively, as a function of the correlation. Here it appears that substantial errors are made in modeling mean system life for even a small amount of dependence.

3. Errors in Estimating Component Parameters

3.1 Parametric Estimation

In this section we examine the effects of incorrectly assuming independent component lifetimes on the magnitude of the estimation error in estimating the first component mean life based on data from series systems. Suppose that n series systems

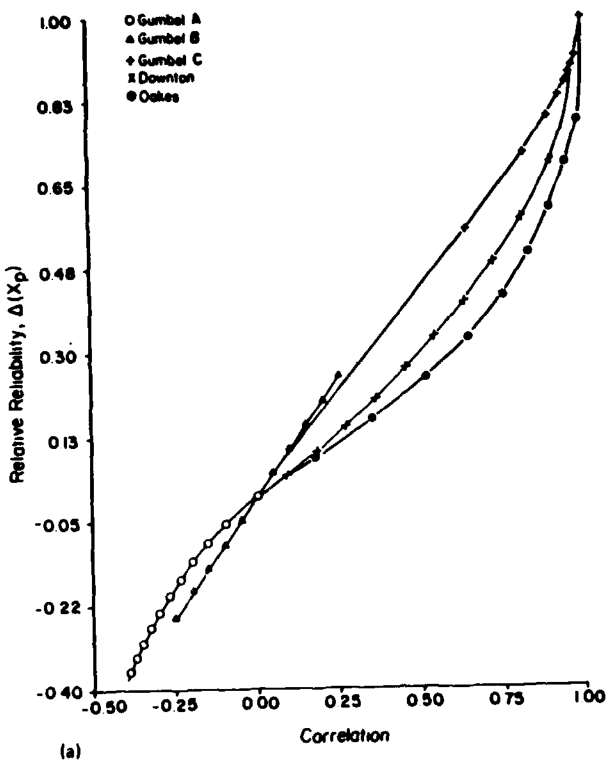


FIGURE 1A
Relative Error in Modeling System Reliability at p =.25
$\lambda_1 = 1.0$, $\lambda_2 = 1.0$

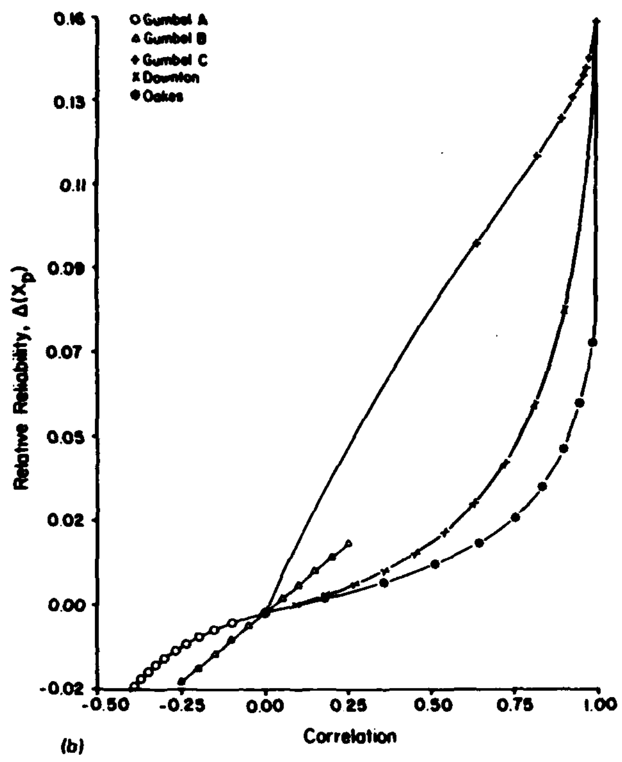


FIGURE 1B
Relative Error in Modeling System Reliability at p =.75
$\lambda_1 = 1.0$, $\lambda_2 = 1.0$

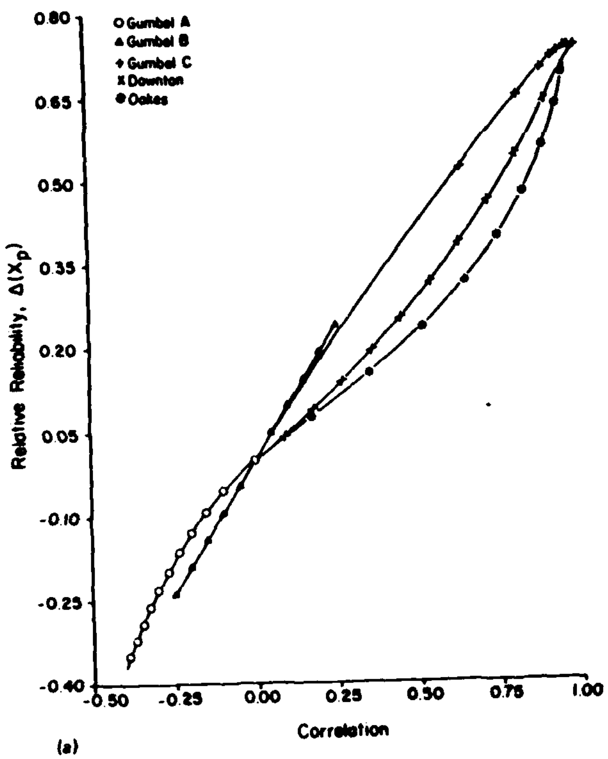


FIGURE 2A

Relative Error in Modeling System Reliability at p =.25

$\lambda_1 = 1.0,\ \lambda_2 = 1.5$

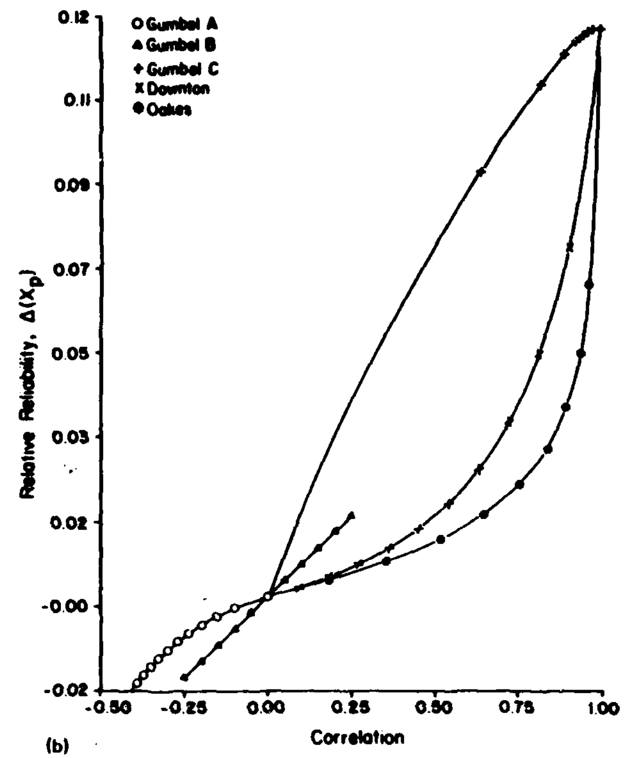


FIGURE 2B

Relative Error in Modeling System Reliability at p =.75

$\lambda_1 = 1.0,\ \lambda_2 = 1.5$

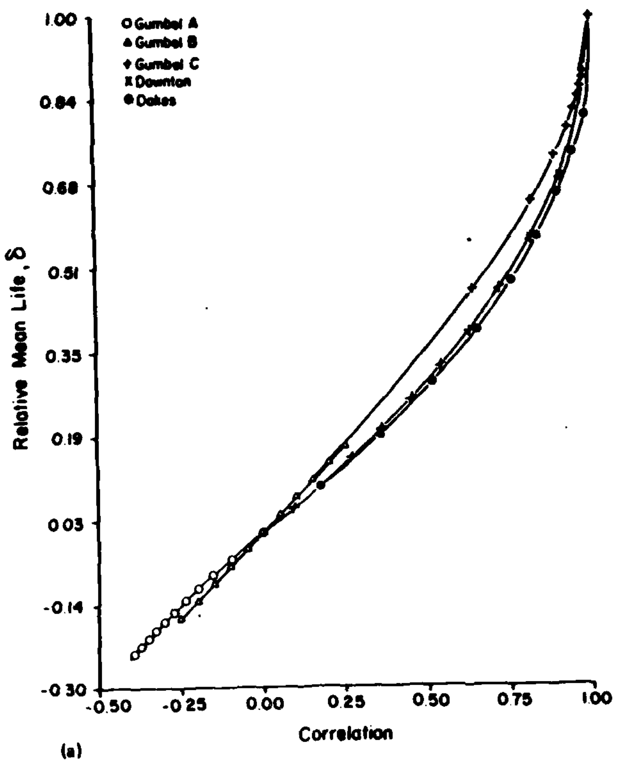


FIGURE 3A
Relative Error in Modeling Mean System Life
$\lambda_1 = 1.0.\ \lambda_2 = 1.0$

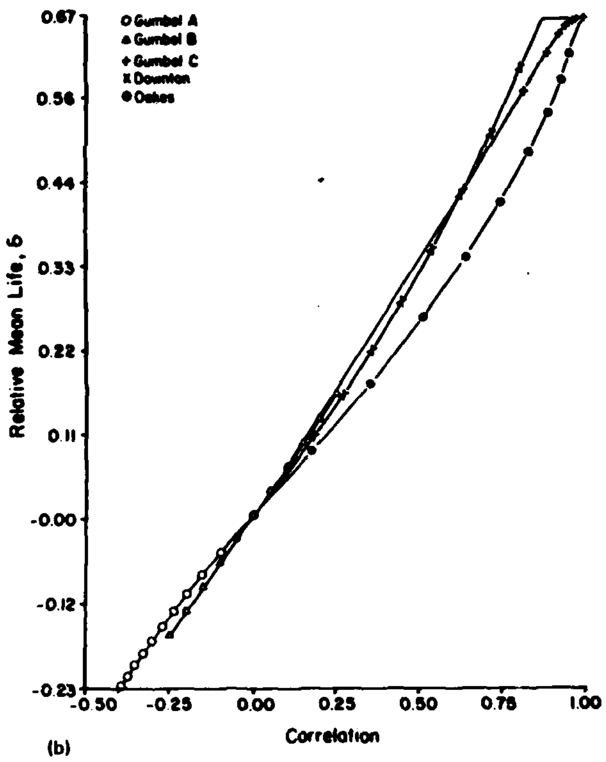


FIGURE 3B
Relative Error in Modeling Mean System Life
$\lambda_1 = 1.0,\ \lambda_2 = 1.5$

are put on test. For each system we observe the system failure time and which component caused the failure. Let n_i denote the number of systems where the system failure was caused by failure of the i^{th} component, $i = 1, 2$, and let T be the total time on test for all n systems. If we assume that the component lifetimes are independent and exponentially distributed then Moeschberger and David (1971) show that the maximum likelihood estimator of μ_1, the first component mean life is

$$\hat{\mu}_1 = T/n_1 \text{ for } n_1 > 0. \tag{3.1}$$

This estimator is asymptotically unbiased and for n finite $E(\hat{\mu}_1) = E(T)\, E(1/n_1 | n_1 > 0)$ due to the independence of T and n_1.

Suppose now that the two component lifetimes are not independent but follow one of the models described in Section 1. If we incorrectly assume independence then a measure of the excess bias due to incorrectly assuming independence is $B = (E(\hat{\mu}_1 | \text{Dependence}) - E(\hat{\mu}_1 | \text{Independence}))/\mu_1$. For each of the dependent models under consideration T and n_1 are independent. For large n, B converges to $(\mu/p - \mu_1)/\mu_1$ where μ is the mean system life and p is the probability the first component fails first, computed under the dependent model. For finite n, $E(\hat{\mu}_1) = n\mu E_p(1/n_1 | n_1 > 0)$ computed under the appropriate model, where

$$E_p(1/n_1 | n_1 > 0) = \sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^k / k \,/\, (1-(1-p)^n).$$

Expressions for μ and p are given in Appendix 1 and Appendix 2, respectively. The expressions depend on λ_1, λ_2 only through the ratio $K = \lambda_1/\lambda_2$. For all models, $p = 1/2$ when $K = 1$.

For the upper Fréchet distribution $p = 0$ if $K < 1$; 1/2 if $K = 1$; and 1 if $K > 1$. Hence for $K < 1$ no failures from the first component are ever observed so that the modeling error B becomes infinite for all n. For $K > 1$, $p = 1$ and $\mu = \mu_1$ so that $B = (1 - E(\hat{\mu}_1 | \text{Independence})/\mu_1)$ which tends to 0 as $n \to \infty$. In this case the models with correlation ranging from 0 to 1 have B increasing for $p < p_0$ and decreasing for $p > p_0$. For the lower Fréchet distribution, p is the value of X which solves the equation $X^K + X - 1 = 0$. For $K < 1$ we have $p < 1/2$ and for $K > 1$ we have $p > 1/2$. Table II gives the value of B for n = 25, 50, ∞ for the two Fréchet distributions. It also gives the maximum modeling error for the Gumbel C model which is an indication of maximal excess modeling error.

From Table II we note that the dependence structure exerts a large effect on estimating the smaller of the two component means and that either effect is most exaggerated for small sample sizes. For $K \geq 1$ there is very little sample size effect on the modeling error. For K strictly larger than one the maximum bias under the Gumbel C model decreases with K and the correlation at which this maximum is attained also decreases to 0.

Figures 4A, 4B are plots of the relative excess bias, B, due to incorrectly assuming independence, as $n \to \infty$ for K = 1.5, and .67, respectively. Figures 5A, 5B are plots of B for n = 10 with K = 1.5, and .67, respectively. For K = 1.5 the sample size effect is negligible in assessing the relative excess bias. For K = .67 the sample size has a noticeable effect on determining B.

3.2 Nonparametric Estimation

A second approach to the problem of estimating component parameters is via the nonparametric estimator of Kaplan and Meier (1958). Investigators who routinely use nonparametric techniques may take this approach in hopes of obtaining estimators that are robust with respect to the assumption of exponentiality. However this estimator is not necessarily robust to the assumption of independence.

The product limit estimator, assuming independent risks is constructed as follows. Suppose that n systems are put on test and let $r_{i1}, \ldots, r_{in_i}$ be the ranks of the ordered n_i failures from cause i, $x_{i\,(1)}, \ldots, x_{i(n_i)}$, among all n

TABLE II
RELATIVE MODELING ERROR IN ESTIMATING THE MEAN

	N=20				N=50				LIMIT (N → ∞)			
	LOWER		GUMBEL C		LOWER		GUMBEL C		LOWER		GUMBEL C	
K	BOUND	RHO=1	RHO	BIAS	BOUND	RHO=1	RHO	BIAS	BOUND	RHO=1	RHO	BIAS
1/10	-167.46	******	1.000	******	-80.09	******	1.000	******	-58.64	******	1.000	******
1/ 9	-142.98	******	1.000	******	-76.35	******	1.000	******	-57.84	******	1.000	******
1/ 8	-123.37	******	1.000	******	-76.68	******	1.000	******	-56.93	******	1.000	******
1/ 7	-107.25	******	1.000	******	-69.03	******	1.000	******	-55.87	******	1.000	******
1/ 6	-93.66	******	1.000	******	-65.34	******	1.000	******	-54.63	******	1.000	******
1/ 5	-81.91	******	1.000	******	-61.54	******	1.000	******	-56.13	******	1.000	******
1/ 4	-71.45	******	1 000	******	-57.50	******	1.000	******	-51.24	******	1.000	******
1/ 3	-61.78	******	1.000	******	-53.00	******	1.000	******	-48.73	******	1.000	******
1/ 2	-52.16	******	1.000	******	-47.31	******	1.000	******	-48.08	******	1.000	******
1	-40.90	105.99	1.000	105.99	-39.45	102.13	1.000	102.13	-38.63	100.00	1.000	100.00
2	-32.56	-2.79	0.510	10.82	-32.84	-1.04	0.518	11.50	-32.12	0.00	0.528	11.91
3	-28.28	-1.82	0.400	5.33	-28.39	-0.69	0.410	5.84	-28.39	0.00	0.424	6.15
4	-25.63	-1.35	0.346	3.39	-25.76	-0.52	0.362	3.76	-25.83	0.00	0.373	4.01
5	-23.62	-1.08	0.312	2.39	-23.80	-0.41	0.330	2.71	-23.90	0.00	0.341	2.91
6	-22.05	-0.89	0.288	1.82	-22.25	-0.34	0.307	2.09	-22.36	0.00	0.319	2.25
7	-20.77	-0.76	0.271	1.46	-20.98	-0.29	0.290	1.68	-21.11	0.00	0.303	1.83
8	-19.69	-0.67	0.257	1.20	-19.92	-0.26	0.277	1.40	-20.04	0.00	0.289	1.53
9	-18.78	-0.59	0.246	1.01	-19.00	-0.23	0.266	1.19	-19.13	0.00	0.279	1.31
10	-17.98	-0.53	0.237	0.87	-18.20	-0.20	0.237	1.03	-18.33	0.00	0.271	1.14

order lifetimes. The estimator of the component reliability for the i^{th} component is

$$\hat{S}_i(x) = \begin{cases} 1 \text{ if} & x < x_{i(1)}, \\ \prod_{j=1}^{j(i,x)} \left(\dfrac{n - r_{ij}}{n - r_{ij} + 1} \right) & x > x_{(i(1))} \end{cases} \tag{3.2.1}$$

where $j(i,x)$ is the largest value of j for which $x_{i(j)} < x$. This estimator is asymptotically unbiased when the component lifetimes are independent.

When the risks are dependent Klein and Moeschberger (1984) show that $\hat{S}_i(x)$ does not estimate the marginal component reliability, but rather it estimates consistently another survival function

$$\bar{H}_i(x) = \exp\left(- \int_0^x \frac{d\,Q_i(t)}{\bar{F}(t)} \right) \tag{3.2.2}$$

where $\bar{F}(x) = P(\text{minimum } (X_1,X_2) > x)$ and $Q_i(x) = P\ (\min(X_1,X_2) \le x,\ \min(X_1,X_2) = X_i),\ i = 1,\ 2.$

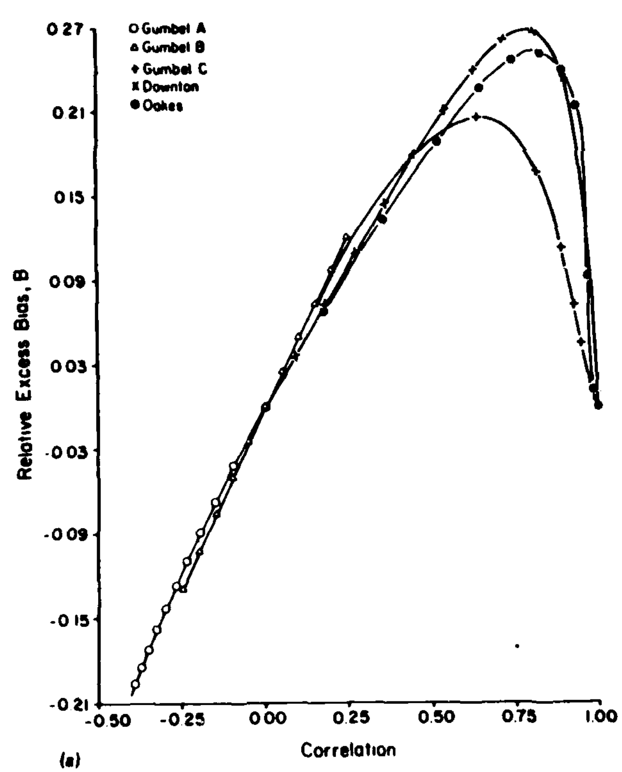


FIGURE 4A
Asymptotic Modeling Error in Estimating μ_1 for K = 1.5

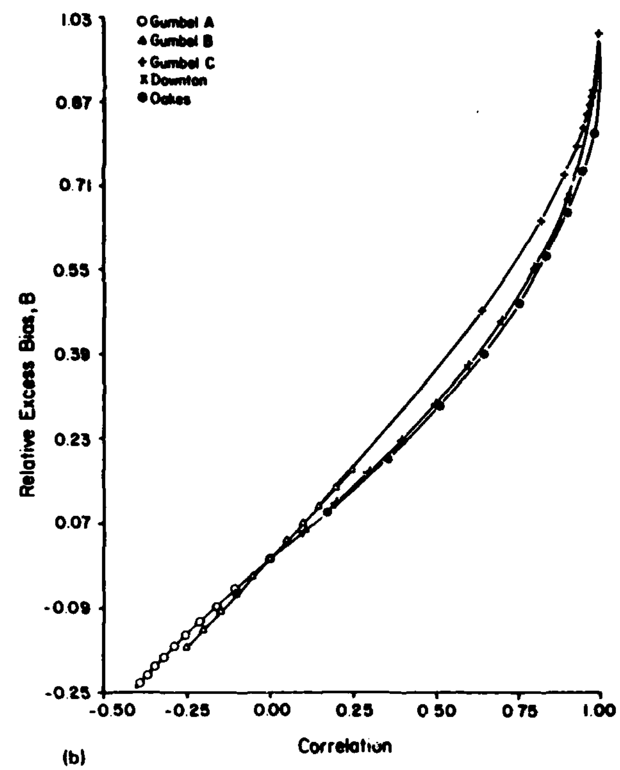


FIGURE 4B
Asymptotic Modeling Error in Estimating μ_1 for K = .67

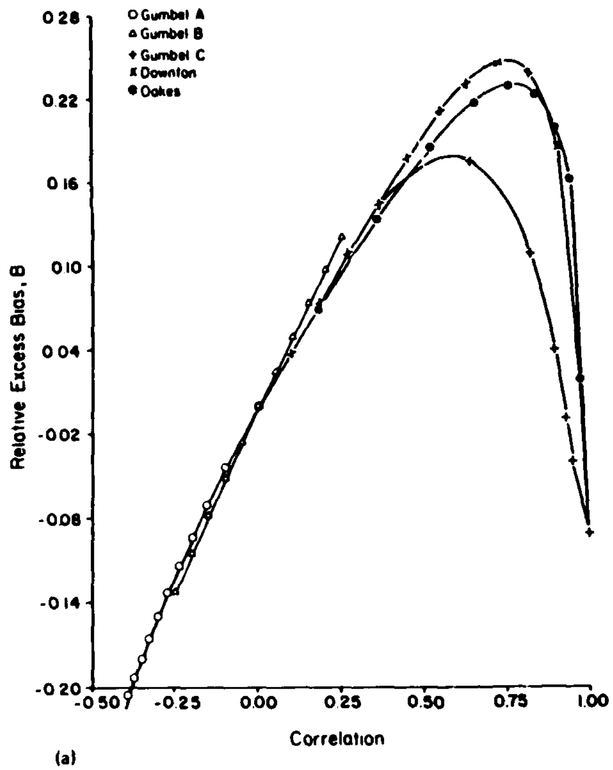


FIGURE 5A
Small Sample Size (N=10) Modeling Error in Estimating
μ_1 for K = 1.5

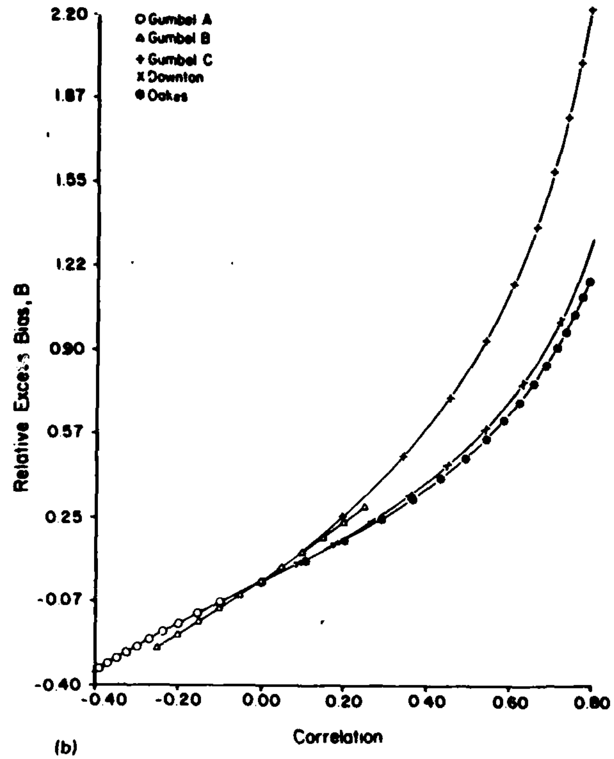


FIGURE 5B
Small Sample Size Modeling Error (N=10) in Estimating
μ_1, K = .67

TABLE III
ASYMPTOTIC BIAS OF THE PRODUCT LIMIT ESTIMATOR

	$\bar{F}(x)=0.7$				$\bar{F}(x)=0.5$				$\bar{F}(x)=0.3$			
	LOWER		GUMBEL C		LOWER		GUMBEL C		LOWER		GUMBEL C	
K	BOUND	RHO=1	RHO	BIAS	BOUND	RHO=1	RHO	BIAS	BOUND	RHO=1	RHO	BIAS
1 /9	-100.00	42.86	1.000	42.86	-100.00	100.00	1.000	100.00	-100.00	233.33	1.000	233.33
1/ 8	-100.00	42.86	1.000	42.86	-100.00	100.00	1.000	100.00	-100.00	233.33	1.000	233.33
1/ 7	-100.00	42.86	1.000	42.86	-100.00	100.00	1.000	100.00	-100.00	233.33	1.000	233.33
1/ 6	-100.00	42.86	1.000	42.86	-100.00	100.00	1.000	100.00	-100.00	233.33	1.000	233.33
1/ 5	-100.00	42.86	1.000	42.86	-100.00	100.00	1.000	100.00	-100.00	233.33	1.000	233.33
1/ 4	-100.00	42.86	1.000	42.86	-100.00	100.00	1.000	100.00	-100.00	233.33	1.000	233.33
1/ 3	-100.00	42.86	1.000	42.86	-100.00	100.00	1.000	100.00	-52.70	233.33	1.000	233.33
1/ 2	-100.00	42.86	1.000	42.86	-100.00	100.00	1.000	100.00	-24.21	233.33	1.000	233.33
1	-100.00	19.52	1.000	19.52	-100.00	41.42	1.000	41.42	-9.65	82.57	1.000	82.57
2	-100.00	0.00	0.528	3.87	-24.67	0.00	0.528	7.65	-4.40	0.00	0.528	13.67
3	-100.00	0.00	0.424	2.09	-14.98	0.00	0.424	4.10	-2.85	0.00	0.424	7.22
4	-63.03	0.00	0.373	1.38	-10.78	0.00	0.373	2.70	-2.11	0.00	0.373	4.74
5	-1.67	0.00	0.341	1.01	-8.42	0.00	0.341	1.98	-44.50	0.00	0.341	3.47
6	-1.39	0.00	0.319	0.79	-6.91	0.00	0.319	1.54	-34.83	0.00	0.319	2.70
7	-1.18	0.00	0.303	0.64	-5.86	0.00	0.303	1.26	-28.71	0.00	0.303	2.19
8	-1.03	0.00	0.290	0.54	-5.09	0.00	0.290	1.06	-24.45	0.00	0.290	1.83
9	-0.92	0.00	0.279	0.46	-4.50	0.00	0.279	0.90	-21.31	0.00	0.279	1.57
10	-0.82	0.00	0.271	0.40	-4.03	0.00	0.271	0.78	-18.89	0.00	0.271	1.37

Expressions for $\bar{H}_1(x)$ for the five models of interest are given in Appendix 3.

A measure of the effect of dependence in using the product limit estimator with dependent risks is $\Delta_1(p) = (\bar{H}_1(x_p) - p)/p$ where x_p is the time where the true component reliability is p. $\Delta_1(p)$ is again only a function of $K = \lambda_1/\lambda_2$. For the upper Fréchet distribution

$$\Delta_1(p) = \begin{cases} p^{-1} - 1 & \text{for} \quad K < 1 \\ p^{-1/2} - 1 & \text{for} \quad K = 1 \\ 0 & \text{for} \quad K > 1 \end{cases}$$

for $K < 1$ $\bar{H}_1(x) = 1$ for all x since the first component never fails, while for $K > 1$ all failures are due to the first component. For those models with correlation spanning the range (0, 1), $\Delta_1(p)$ is increasing for correlations less than p and decreasing for correlations greater than p when $K > 1$. For the lower Fréchet distribution

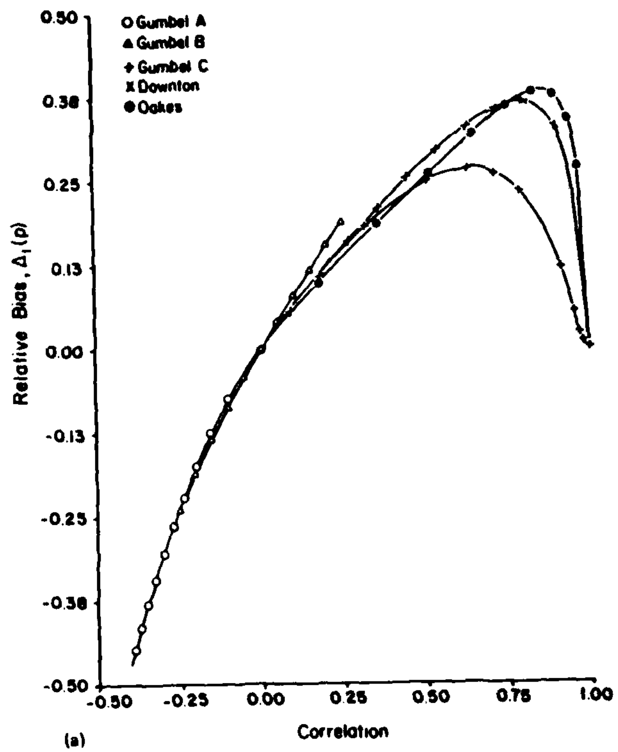


FIGURE 6A
Asymptotic Error in the Nonparametric Estimator of the First Component Survival Function at
p = .25, K = 1.5

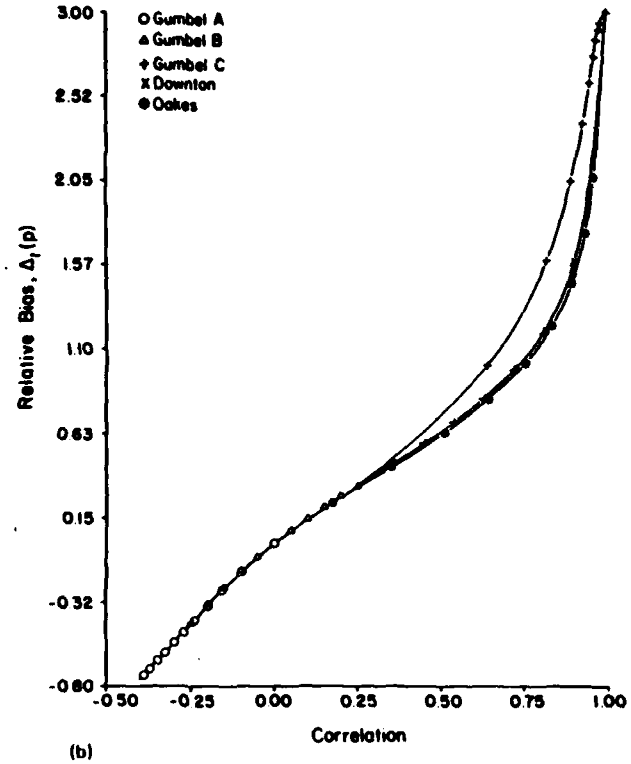


FIGURE 6B
Asymptotic Error in the Nonparametric Estimator of the First Component Survival Function at
p = .25, K = .67

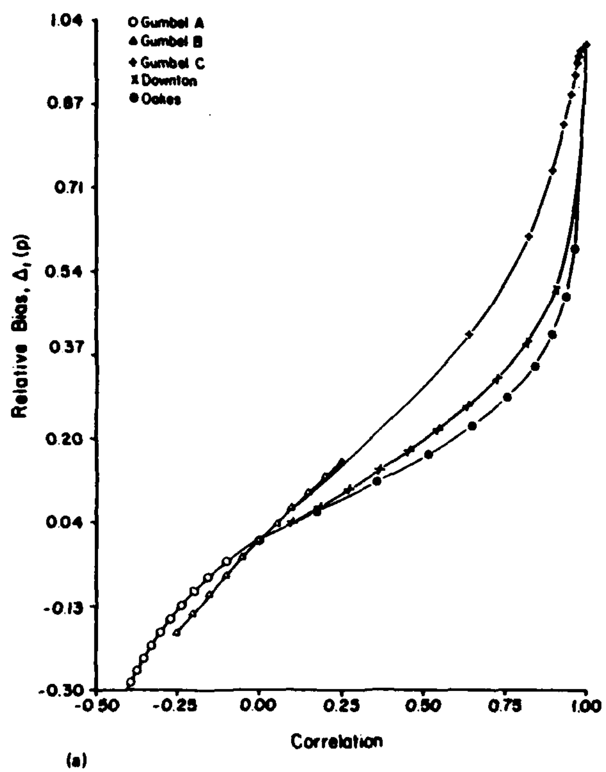


FIGURE 7A
Asymptotic Error in the Nonparametric Estimator of the First Component Survival Function at
p = .5, K = 1.5

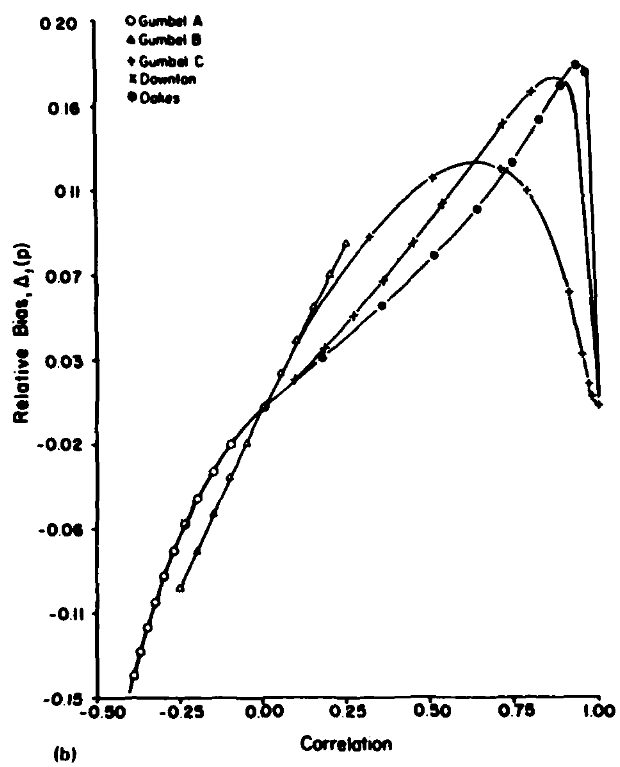


FIGURE 7B
Asymptotic Error in the Nonparametric Estimator of the First Component Survival Function at
p = .5, K = .67

$$\bar{H}_1(x_p) = \begin{cases} \exp\left[-\int_p^1 \left(\dfrac{1}{u + u^k - 1} \right) du \right] & \text{for } p \geq (1-Y) \\ 0 & \text{otherwise,} \end{cases}$$

where Y is the solution to the equation $X^K + X = 1$. Table III shows the value of $\Delta_1(p) \times 100\%$ for $p = .7, .5, .3$ for the two Fréchet distributions. For $K > 1$, the maximum value under the Gumbel C model is also given. As in the parametric estimation problem the largest errors are incurred when $K < 1$. In all cases the effect of a departure from independence is the largest when p is small (i.e. for large x). The effect decreases as K increases reflecting the fact that when $K\lambda_1 >> \lambda_2$ the majority of the system failures are due to the failure of the first component.

Figures 6A, 6B for $p = .25$ and 7A, 7B for $p = .5$ are plots of $\Delta_1(p)$ for the five models and $K = 1.5, .67$, respectively. As in the previous figures one can see that for even a small departure from independence the relative effect of dependence can be quite large.

4. Conclusions

The results presented in this paper show that for all five bivariate exponential models one may be appreciably misled by falsely assuming independence of component lifetimes in a series system. The amount of error incurred in modeling system reliability not only depends upon the correlation between component lifetimes but also on the level of system reliability. The error in modeling mean system life similarly depends upon the correlation and the length of mean system life. Both quantities depend on the relative magnitudes of the parameters.

For the dual problem of estimating component reliability based on data from a series system, it appears that departures from independence are of greater consequence. Both parametric and nonparametric estimators of relevant component parameters are inconsistent. Bias increases dramatically as the correlation gets further from zero.

ACKNOWLEDGMENT

Research sponsored by the Air Force Office of Scientific Research under contract AFOSR-82-0307

BIBLIOGRAPHY

Barlow, R.E. and Proschan, F. (1975). Statistical Theory of Reliability and Life Testing. Holt Rinehart, and Winston, Inc., New York.

Basu, A.P. (1981). Identifiability problems in the theory of competing and complementary risks - a survey. Statistical Distributions in Scientific Work (Taillie, Patil, and Baldesaari, Eds.), Dorerecht, Holland: Reidel Publishing Co., 335-348.

Basu, A.P. and Klein, J.P. (1982). Some Recent Results in Competing Risks Theory. Survival Analysis, (Crowley and Johnson, Eds.), Institute of Mathematical Statistics, Hayward, CA, 216-229.

Birnbaum, Z.W. (1979). On the Mathematics of Competing Risks. U.S. Dept. of HEW. Publication No. 79-1351.

Boardman, T.J. and Kendell, P.J. (1970). Estimation in compound exponential failure models. Technometrics 12, 891-900.

Clayton, D.G. (1978). A model for association in bivariate lifetables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65, 141-151.

David, H.A. and Moeschberger, M.L. (1978) The Theory of Competing Risks. Griffin, London.

Downton, F. (1970). Bivariate exponential distributions in reliability theory. J. Roy. Statist. Soc. B. 32: 408-417.

Easterling, R.G. (1980). Book review of Z.W. Birnbaum monograph. Technometrics 22, 131-132.

Frechet, M. (1958). Sur les Tableaux de correlation Dont les Marges Sont Donnees. Annales de l'Universite de Lyon. Section A. Series 3, 14, 53-77.

Gumbel, E.J. (1960). Bivariate exponential distributions. J. Amer. Statist. Assoc. 55, 698-707.

Hutchinson, T.P. (1981). Compound gamma bivariate distributions. Metrika, 28, 263-271.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457-481.

Klein, J.P. and Moeschberger, M.L. (1984). Asymptotic bias of the product limit estimator under dependent competing risks. IAPQR Transactions, 9, 1-7.

Lagakos, S.W. (1979). General right censoring and its impact on the analysis of survival data. Biometrics 35, 139-156.

Lagakos, S.W. and Williams, J.S. (1978). Models for censored survival analysis: a cone class of variable-sum models, Biometrika, 65, 181-189.

Lindley, D. and Singpurwalla, N. (1985). Multivariate distributions for the reliability of a system of components sharing a common environment. Technical Report. George Washington University. Washington, D.C.

Mann, N.R. and Grubbs, F.E. (1974). Approximately optimum confidence bounds for system reliability based on component test data. Technometrics, 16, 335-347.

Mann, N.R., Schafer, R.E., and Singpurwalla, N.D. (1974). Methods for the Statistical Analysis of Reliability and Life Data. Wiley, New York.

Miller, D.R. (1977). A note on independence of multivariate lifetimes in competing risks. Ann. Statistics. 5: 576-579.

Miyamura, T. (1982). Estimating component failure rates from combined component and systems data: exponentially distributed component lifetimes. Technometrics, 24: 313-318.

Moeschberger, M.L. and David, H.A. (1971). Life tests under competing causes of failure and the theory of competing risks. Biometrics 27, 909-933.

Moeschberger, M.L. and Klein, J.P. (1984). Consequences of departures from independence in exponential series systems. Technometrics 26: 277-284.

Oakes, D. (1982). A model for association in bivariate survival data. J. Roy. Statist. Soc. B. 44: 414-422.

Peterson, A.V., Jr. (1976). Bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks. Proc. Nat. Acad. Sci. 73: 11-13.

Williams, J.S. and Lagakos, S.W. (1977). Models for censored survival analysis: Constant-sum and variable-sum models. Biometrika 64: 215-224.

APPENDIX 1

Formulas for expected system life.

Gumbel A: $$\exp \frac{(\lambda_1+\lambda_2)^2}{4\lambda_1 \lambda_2} \sqrt{\frac{\Pi}{\lambda_{12}}}\ \bar{\Phi}\left(\frac{\lambda_1+\lambda_2}{\sqrt{2}\ \lambda_{12}}\right), \tag{A1.1}$$

where $\bar{\Phi}(\cdot)$ is the survival function of a standard normal random variable.

Gumbel B: $$\frac{1}{(\lambda_1+\lambda_2)} + \frac{6\ \rho\ \lambda_1\lambda_2}{(\lambda_1+\lambda_2)(2\lambda_1+\lambda_2)(\lambda_1+2\lambda_2)} \tag{A1.2}$$

Gumbel C: $$(\lambda_1^m+\lambda_2^m)^{-1/m} \tag{A1.3}$$

Downton : $$\frac{(\lambda_1+\lambda_2)(1-\rho)}{((\lambda_1+\lambda_2)^2 - 4\rho\ \lambda_1\lambda_2)} + \frac{\left[(\lambda_1+\lambda_2) - \sqrt{(\lambda_1+\lambda_2)^2 - 4\rho\ \lambda_1\lambda_2}\right](\lambda_1^2+\lambda_2^2 - 2\rho\ \lambda_1\lambda_2)}{2\lambda_1\lambda_2((\lambda_1+\lambda_2)^2 - 4\rho\ \lambda_1\lambda_2)} \tag{A1.4}$$

Oakes : found by numerical integration

APPENDIX 2

Formulas for $p = P(X_1 < X_2)$:

Gumbel A - $P(X_1 < X_2)$

$$= 1/2 + (\lambda_1-\lambda_2)\sqrt{\frac{\Pi}{\lambda_{12}}}\ \exp\left(\frac{(\lambda_1+\lambda_2)^2}{4\lambda_{12}}\right)\bar{\Phi}\left(\frac{\lambda_1+\lambda_2}{2\lambda_{12}}\right),\quad \lambda_{12} > 0$$

where $\bar{\Phi}(\cdot)$ is the survival function of a standard normal random variable. (A2.1)

Gumbel B - $P(X_1 < X_2)$

$$= \frac{\lambda_1}{(\lambda_1+\lambda_2)} + \frac{4\rho\ \lambda_2(\lambda_1-\lambda_2)}{(\lambda_1+\lambda_2)(\lambda_1 + 2\lambda_2)(2\lambda_1 + \lambda_2)} \tag{A2.2}$$

Gumbel C - $P(X_1 < X_2)$

$$= \frac{\lambda_1^m}{(\lambda_1^m + \lambda_2^m)} \tag{A2.3}$$

Downton - $P(X_1 < X_2)$

$$= \frac{2\lambda_1\lambda_2(1-\rho)}{\sqrt{((\lambda_1+\lambda_2)^2 - 4\rho\,\lambda_1\lambda_2)(\lambda_1-\lambda_2)} + \sqrt{(\lambda_1+\lambda_2)^2 - 4\rho\lambda_1\lambda_2}} \tag{A2.4}$$

Oakes - $P(X_1 < X_2)$ found numerically.

APPENDIX 3

Gumbel A - $\bar{H}_1(x) = \exp(-\lambda_1 x - \lambda_{12} x^2)$ (A3.1)

Gumbel B - $$\bar{H}_1(x) = \exp\left\{-\int_0^x \frac{(1 + 4\rho(1-\exp(-\lambda_2 t))(1-2\exp(-\lambda_1 t)))}{(1 + 4\rho(1-\exp(-\lambda_2 t))(1-\exp(-\lambda_1 t)))}\,dt\right\} \tag{A3.2}$$

Gumbel C - $$\bar{H}_1(x) = \exp\left(-\frac{\lambda_1^m x}{(\lambda_1^m + \lambda_2^m)^{\frac{m-1}{m}}}\right) \tag{A3.3}$$

Downton - Found numerically due to no closed form solution for $\bar{F}(x)$.

Oakes - $$\bar{H}_1(x) = \exp\left\{-\int_0^x \frac{\lambda_1 \exp(\lambda_1(\theta-1)t)}{(\exp(\lambda_1(\theta-1)t) + \exp(\lambda_2(\theta-1)t) - 1)}\,dt\right\} \tag{A3.4}$$

Received by Editorial Board member Ausut, 1985; Revised and retyped December, 1986.

Recommended by R. L. Scheaffer, University of Florida; Gainesville, FL.

Refereed by James Williams, Colorado State University, Fort Collins, CO.

communications in statistics

SIMULATION AND COMPUTATION

THEORY AND METHODS

STOCHASTIC MODELS

ORDER FORM

COMMUNICATIONS IN STATISTICS
Part A: *Theory and Methods; Vol. 16, 12 issues*
Part B: *Simulation and Computation; Vol. 16, 4 issues*
Part C: *Stochastic Models; Vol. 3, 3 issues*

COMBINATION RATE FOR 19 ISSUES
Institutional: $827.00
Individual: $413.50

(Please add $57.00 for foreign surface postage, For airmail to Europe, add $71.00; to Asia, add $90.00.)

All subscriptions must be prepaid. Payment must be made in U.S. Currency.

Please enter my subscription for:

COMMUNICATIONS IN STATISTICS

at the ☐ institutional ☐ individual rate.

I enclose payment by:

☐ **Check** ☐ **Money Order** **in the amount of $** ______________

☐ **Visa** ☐ **Master Card No.** ______________ **Exp. Date** ______________

Master Card four digit interbank number ______________

Signature ______________________________
(must accompany credit card payment)

Name ______________________________

Address ______________________________

City ______________________________

State ______________________ **Zip** ______________

☐ **Please send me a complimentary copy of this journal.**

Please recommend library acquisition of this and all materials needed for teaching and research.

MAIL TO: **PROMOTION DEPARTMENT**
MARCEL DEKKER, INC.
270 MADISON AVENUE
NEW YORK, NEW YORK 10016

CISAB-R86

APPENDIX E

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY NA	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE NA	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION The Ohio State University	6b. OFFICE SYMBOL (If applicable)
7a. NAME OF MONITORING ORGANIZATION AFOSR/NM	
6c. ADDRESS (City, State, and ZIP Code) 1314 Kinnear Road Columbus, Ohio 43212	7b. ADDRESS (City, State, and ZIP Code) AFOSR/NM Bldg 410 Bolling AFB DC 20332-6448
8a. NAME OF FUNDING / SPONSORING ORGANIZATION AFOSR	8b. OFFICE SYMBOL (If applicable) NM
9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-82-0307	
8c. ADDRESS (City, State, and ZIP Code) AFOSR/NM Bldg 410 Bolling AFB DC 20332-6448	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
61102F	2304		

11. TITLE (Include Security Classification) Bounds on Net Survival Probabilities For Dependent Competing Risks

12. PERSONAL AUTHOR(S) John P. Klein and M. L. Moeschberger

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 9-1-82 TO 12-31-87	May 31, 1988	19

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP
XXXX	XXXXXXXXXX	XXX

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Competing risks, product limit estimator, net survival function, coefficient of concordance

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Bounds on the marginal survival function based on data from a competing risks experiment are obtained. These bounds require an investigator to specify a range of possible concordances for the times to occurrences of the competing risks. These bounds are tighter than those of Peterson (1976).

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Maj. Brian Woodruff	22b. TELEPHONE (Include Area Code) (202) 767-5027
22c. OFFICE SYMBOL NM	

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted

All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

BOUNDS ON NET SURVIVAL PROBABILITIES
FOR DEPENDENT COMPETING RISKS

by

John P. Klein[1] and M.L. Moeschberger[2]

Departments of Statistics[1] and Preventive Medicine[2],
The Ohio State University, 1958 Neil Avenue,
Columbus, Ohio 43210, U.S.A.

SUMMARY

Bounds on the marginal survival function based on data from a competing risks experiment are obtained. These bounds require an investigator to specify a range of possible concordances for the times to occurrences of the competing risks. These bounds are tighter than those of Peterson (1976).

Key Words: Competing Risks, Product Limit Estimator, Net Survival Function, Coefficient of Concordance.

1. Introduction

A common problem encountered in biological and medical studies (both animal and human) is to estimate the marginal survival function of the time X, from some appropriate starting point, until some event of interest (such as the occurrence of a particular disease, remission, relapse, death due to some specific disease or simply death) occurs. Often it is impossible to measure X due to the occurrence of some other competing event, usually termed a competing risk, at some time $Y(< X)$. This competing event may be the withdrawal of the subject from the study (for whatever reason), death from some cause other than the one of interest, or any eventuality which precludes the main event of interest from occurring. In such instances the actual time until the main event of interest occurs can be regarded as censored (see David and Moeschberger (1978) for further discussion of such censoring). With such a competing risks representation, it is often assumed that this censoring time is independent of the main event of interest. This allows for tl consistent estimation of the marginal distribution of X, $S(x) = P[X \geq x]$. This assumption of independence is also made in more complex analyses of competing risks experiments such as the use of log-linear models for the analysis of survival data (Holford (1980)), the analysis of proportional hazards regression of censored data (Cox (1972)), computation of Hodges-Lehmann like estimators with censored data (Wei and Gail (1983)), to name a few.

A standard statistical estimator of the survival function which assumes such competing events (or risks) to be independent is the Product Limit Estimator of Kaplan and Meier (1958). This estimator is nonparametric and consistent for the class of constant sum survival models defined by Williams and Lagakos (1977). When the risks are not in this class the Product Limit Estimator is inconsistent and, in such cases, the investigator may be appreciably misled by assuming independence (see Lagakos (1979), Lagakos and Williams (1978), Moeschberger and Klein (1984) for details).

In the competing risks framework we observe T = minimum (X, Y) and I = $\chi(X \leq Y)$ where $\chi(\cdot)$ denotes the indicator function. It is well known (see Basu and Klein (1982) for details and references) that the pair (T,I) provide insufficient information to determine the joint distribution of X and Y. That is, there exists both an independent and one or more dependent models for (X,Y) which produces the same joint distribution for (T,I). However, these "equivalent" independent and dependent joint distributions may have quite different marginal distributions. In light of the consequences of the untestable independence assumption in using the Product Limit Estimator to estimate the marginal survival function of X, it is important to have bounds on this function based on the observable random variables (T,I) and some assumptions on the joint behavior of X and Y.

Peterson (1976) has obtained general bounds on the marginal survival function of X, S(x), based on the minimal and maximal dependence structure for (X,Y) obtained by Fréchet (1951). Let $P_1(x) = P(T \geq x, I=1)$ and $P_2(x) = P(T > x, I=0)$ be the crude survival functions of T. The bounds are $P_1(x) + P_2(x) \leq S(x) \leq P_1(x) + P_2(0)$.

These bounds allow for any possible dependence structure and can be very wide.

Slud and Rubenstein (1983) have obtained tighter bounds on $S(x)$ in this framework by utilizing some additional information. Their method requires the *investigator to* bound the function

$$\rho(t) = [\{s(t)/q_1(t)\} - 1] \big| [\{S(t)/F(t)\} - 1], \tag{1.1}$$

where

$s(t) = - d\, S(t)/d\, t$, $F(t) = P(\min(X,Y) > t)$ and

$q_1(t) = \frac{d}{dt} P(T < t,\ X < Y)$. Knowledge of the function, $\rho(t)$, and the observable information, (T,I), is sufficient to uniquely determine the marginal distribution of X. The resulting estimators $\hat{S}_\rho(t)$ are decreasing functions of $\rho(\cdot)$. The resulting bounds are obtained by the investigator's specification of two functions, $\rho_i(x)$, $(\rho_1(x) < \rho_2(x))$ so that if the true $\rho(x)$ function is in the interval $[\rho_1(x),\ \rho_2(x)]$, for all x, then $\hat{S}_{\rho_2}(x) \le S(x) \le \hat{S}_{\rho_1}(x)$.

In the sequel we obtain alternative bounds on the marginal survival function utilizing slightly different additional information. We assume that the joint distribution of the time until death and censoring, (X,Y), belongs to a family of distributions indexed by a dependence measure θ with arbitrary marginals. For this family, knowledge of θ, along with the observable information, (T,I), is sufficient to uniquely determine the marginal distributions of X and Y. The resulting estimator $\hat{S}_\theta(t)$ is a decreasing function of θ so that bounds on $S(t)$ for the family of joint distributions is obtained by specifying a range of possible values for θ.

2. The Model

The dependence structure we shall employ to model the joint distribution of time until death and censoring time was first introduced by Clayton (1978) to model association in bivariate lifetables, and, later, by Oakes (1982) to model bivariate survival data. A revision of this model, with an underlying exponential structure, has been proposed by Lindley and Sigpurwalla (1985) and Hutchinson (1981) as a model for system reliability in an engineering context.

Let $S(x) = P(X \geq x)$ be the univariate survival function of death and $R(y) = P(Y \geq y)$ be the probability of not being censored before time y ($S(0) = R(0) = 1$, S and R continuous functions). For $\theta \geq 1$, define $F(x,y) = P(X > x,\ Y > y)$ by

$$F(x,y) = \left[\left\{\frac{1}{S(x)}\right\}^{\theta-1} + \left\{\frac{1}{R(y)}\right\}^{\theta-1} - 1\right]^{-1/(\theta-1)} \qquad (2.1)$$

This joint distribution has marginals S and R. As $\theta \to 1$, then (2.1) reduces to the joint distribution with independent marginals. For $\theta \to \infty$, $F(x,y) \to \min(S(x), R(y))$ the bivariate distribution with maximal positive association for these marginals. If (X_i, Y_i) and (X_j, Y_j) are independent bivariate random variables with survival function (2.1) then the probability of concordance is $P[(X_i - X_j)(Y_i - Y_j) > 0] = \theta/(\theta + 1)$

so that Kendall's (1962) coefficient of concordance is $\tau = (\theta - 1)/(\theta + 1)$ which spans the range 0 to 1.

The model (2.1) can also be derived from the following random environmental effects model. Let X_0 and Y_0 denote the potential time to failure from the two risks, say death from a particular cause and death from some other cause or causes. Suppose that in a perfect environment X_0 and Y_0 are independent with survival functions $S_0(x) = \exp(-S(x)^{-\theta+1} + 1)$ and $R_0(y) = \exp(-R(y)^{-\theta+1} + 1)$. The individual lives in an environment where various environmental stresses or biological exposures may produce a random effect W which, in turn, changes the potential times to occurrences, X_0 and Y_0, to X and Y with survival functions $S_0^w(x)$ and $R_0^w(y)$, respectively. A value of w less than one implies a joint improvement in the survival probabilities for the two risks, while a value of w greater than one implies a joint degradation. If W has a gamma distribution with density $g(w) \propto w^{\left(\frac{1}{\theta-1}\right)-1} e^{-w}$ then the unconditional distribution of

(X,Y) follows the form (2.1). The stresses or exposures which produce this random effect w may be different biological exposures that are characteristic of the individual's behavior, such as smoking or they may consist, in part, by exposure to an environmental agent, such as asbestos.

This model also has a physical interpretation in terms of the functions $\lambda(x|Y = y) = \lim_{\Delta x \to o} \left\{\frac{\Pr(x \le X < x+\Delta x | Y=y,\ X \ge x)}{\Delta x}\right\}$

and $\lambda(x|Y > y) = \lim_{\Delta x \to o} \left\{\frac{\Pr(x \le X < x+\Delta x | Y>y,\ X \ge x)}{\Delta x}\right\}$, the hazard functions

of X given Y = y and X given Y > y, respectively. These hazard functions specify the instantaneous rate of death or failure at time x, given that the individual is censored at time y or later than y, respectively. From (2.1) one can see that

$$\lambda(x|Y = y) = \theta\lambda(x|Y > y)$$

or

$$P(X > x|Y = y) = [P(X > x|Y > y)]^{\theta} \tag{2.2}$$

For $\theta > 1$ the hazard rate of death, if censoring occurs at time y, is θ times the hazard rate of death, if censoring does not occur at time y. This implies that the hazard rate, after censoring occurs, is accelerated by a factor of θ over the hazard rate if censoring had not occurred. Also when $\theta = 1$, (2.2) reduces to the condition required by Williams and Lagakos (1977) for a model to be constant sum and hence for the usual product limit estimator of S(t) to be consistent (See Basu and Klein (1982) for details).

For fixed marginals S and R the joint probability density function, f(x,y), can be shown to be totally positive of order 2 for all $\theta \geq 1$. This implies that (X,Y) are positive quadrant dependent. In particular, one can show that, for S, R fixed, the family of distributions $F = \{F(x,y)\colon \theta \geq 1\}$ is increasing positive quadrant dependent in θ as defined by Ahmed, et al. (1979).

3. Bounds on Marginal Survival

Suppose that X and Y have the joint distribution (2.1) and let $T = \min(X,Y)$, then the survival function of T is

$$F(t) = [[\frac{1}{S(t)}]^{\theta-1} + [\frac{1}{R(t)}]^{\theta-1} - 1]^{-\frac{1}{\theta-1}} \tag{3.1}$$

and the crude density function associated with X,

$$q_1(t) = \frac{d}{dt} P(T < t, X < Y), \text{ is given by}$$

$$q_1(t) = \frac{s(t)}{S^\theta(t)} [F(t)]^\theta, \tag{3.2}$$

where $s(t) = -dS(t)/dt$.

Now consider the differential equation

$$s(t)/S^\theta(t) = q_1(t)/[F(t)]^\theta \tag{3.3}$$

and suppose θ is known. Then the solution of (3.3) for S(t) is

$$\begin{aligned} S_\theta(t) &= \left[1 + (\theta-1) \int_0^t \frac{q_1(u)}{[F(u)]^\theta} du\right]^{-\frac{1}{(\theta-1)}} && \text{if } \theta > 1 \\ &= \exp\left(- \int_0^t \frac{q_1(u)}{F(u)} du\right) && \text{if } \theta = 1. \end{aligned} \tag{3.4}$$

The functions $F(\cdot)$ and $q_1(\cdot)$ are directly estimable from the data one sees in a competing risks experiment. Let $T_1, \ldots, T_n$ denote the observed test times of n individuals put on test and let I_i, $i = 1, \ldots, n$ be 1 or 0 according to whether the T_i was an observation on X_i or Y_i, respectively.

Define $\hat{F}(t) = \sum_{i=1}^{n} \chi(T_i \geq t)/n$ and $\hat{Q}_1(t) = \sum_{i=1}^{n} \chi(T_i \leq t,\ I_i = 1)/n$.

Then if θ is know, a natural estimator of $S_\theta(t)$ is

$$\hat{S}_\theta(t) = \begin{cases} \left[1 + (\theta-1) \int_0^t \dfrac{d\hat{Q}_1(u)}{[\hat{F}(u)]^\theta}\right]^{-\frac{1}{(\theta-1)}} & \text{if } \theta > 1 \\ \exp\left(-\int_0^t \dfrac{d\hat{Q}_1(u)}{\hat{F}(u)}\right) & \text{if } \theta = 1 \end{cases} \qquad (3.5)$$

For $\theta = 1$, this estimator is of the form of the hazard rate estimator proposed by Nelson (1972). The estimators (3.5) can be expressed in the following forms for computation purposes when there are no ties.

$$\hat{S}_\theta(t) = \begin{cases} \left[1 + (\theta-1)n^{\theta-1} \sum_{\substack{T_{(i)} \leq t \\ I_{(i)}=1}} \dfrac{1}{(n-i+1)^\theta}\right]^{-\frac{1}{\theta-1}} & \text{if } \theta > 1 \\ \exp\left\{-\sum_{\substack{T_{(i)} \leq t \\ I_{(i)}=1}} \dfrac{1}{(n-i+1)}\right\} & \text{if } \theta = 1 \end{cases} \qquad (3.6)$$

where $T_{(1)}, \ldots, T_{(n)}$ are the ordered death times or censoring times.

For θ known and if the true underlying joint distribution of (X,Y) is of the form (2.1) then $\hat{S}_\theta(t)$ is a consistent estimator of $S(t)$ as shown in the Appendix.

To obtain bounds on the net survival function based on data from a competing risks experiment, we proceed as follows. First, note that from (3.4) it is true that $S_\theta(t)$ is a decreasing function of θ for fixed t.

Also, as $\theta \to 1^+$ we have $S_\theta(t) \uparrow \exp\left(-\int_0^t F^{-1}(u)dQ_1(u)\right)$

which provides an upper bound on S(t). Notice that this upper bound corresponds to an assumption of independence. As $\theta \to \infty$ one can show that $S_\theta(t) \downarrow F(t)$ which corresponds to Peterson's (1976) lower bound.

In practice the above bounds, with $\theta = 1, \infty$, while shorter than Peterson's bounds may still be quite wide. Tighter bounds, in the spirit of Slud and Rubenstein, may be obtained if an investigator can specify a range of possible values for θ, (θ_1, θ_2). If the sample size is sufficiently large and $\theta_1 \leq \theta \leq \theta_2$, then $\hat{S}_{\theta_1}(t) \geq S(t) \geq \hat{S}_{\theta_2}(t)$. Two approaches to specifying θ_1, θ_2 are appropriate. From (2.2) note that $\theta = \lambda(x|Y=y)/\lambda(x|Y>y)$ for all x, y, so that θ_1 and θ_2 are reflections of the investigators belief in how the hazard rate of X would be effected by knowledge of the occurrence of censoring at time y. Secondly, specification of θ_1, θ_2 is equivalent to specifying a range of values for the coefficient of concordance, τ, between failure time X and censoring time Y, since $\theta = (1 + \tau)/(1 - \tau)$.

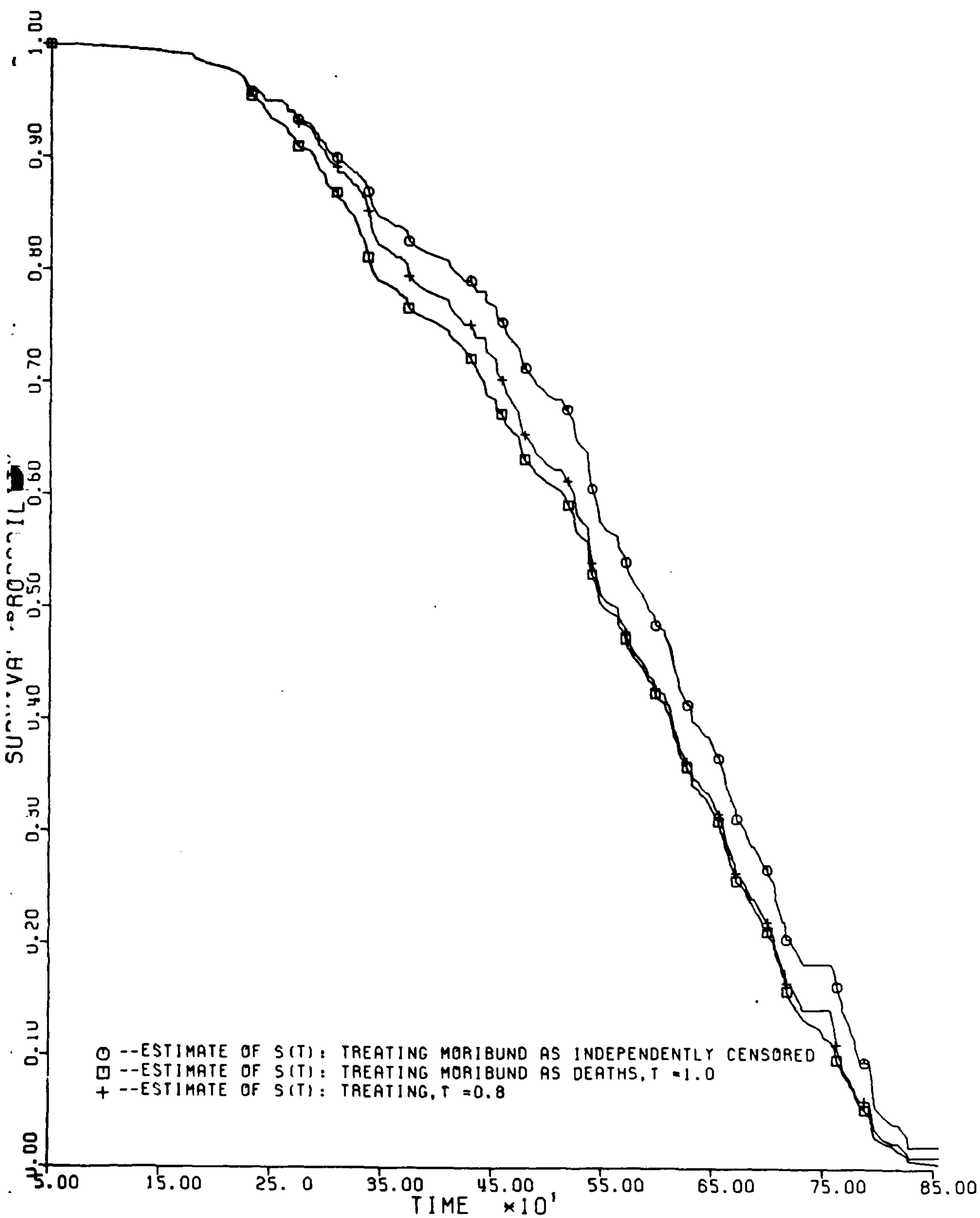
4. Examples and Discussion

To illustrate the bounds obtained in the previous section consider an experiment performed at the Oak Ridge National Laboratory. The experiment we consider consists of treating a group of 246 RFM female mice with 75 rads of gamma radiation delivered at 45 rads/min. at 12 weeks of age. Mice which were judged by the biologist to be moribund and not likely to survive until the next observation period (either overnight or over the weekend) were sacrificed. This action, which caused mice to be removed from the study before the endpoint of death was observed, was necessitated by the need to determine by necropsy which of several types of tumors the animals had at death since the status of such tumors was difficult to determine in animals which lay dead in their cages for an extended period of time. The value of $\tau=1$ ($\theta=\infty$) assumes that the moribund animals would die at a time coincidental with the sacrifice time. It should be noted that the data was actually analyzed under this assumption. Animals which were sacrificed prior to a weekend were probably more likely to live an additional time due to the conservative nature of the decision making. It is clear that the sacrifice (censoring) time and death time of these animals are highly associated, the value of τ probably being somewhat less than one.

A lower bound of τ, though somewhat subjective, may be based on the investigator's experience with such experiments where moribund mice were not sacrificed. In this instance, τ of .8 seems reasonable. The value of $\hat{S}_\theta(t)$ for τ = .8 and 1.0 as well as the product limit estimator assuming independence (a statistical procedure which may also be used by some people) is depicted in Figure 1. One can see that the bounds for the survival function (assuming $.8 \leq \tau \leq 1.0$) provide a relatively tight band of survival functions. It is also clear that the product limit estimator assuming censoring times independent of death times is probably not appropriate.

Consider a second study which illustrates a different type of censoring mechanism. A clinical trial was conducted at The Ohio State University to determine the objective response rate of patients with refractory advanced non-Hodgkin's lymphoma to a chemotherapy regimen consisting of ifosfamide, VP-16, cis-platinum, and bleomycin. Twenty-four patients were entered (staggered entry) and treated with ifosfamide 750

FIGURE 1
ESTIMATES OF SURVIVAL FOR RFM/UM MICE



mg/m2/day by continuous I.V. infusion on days 1-5, VP-16 40 mg/m2 I.V. on days 1-5, cis-platinum 50 mg/m2 I.V. on day 14, and bleomycin 15 units/m2 I.M. on day 14. The regimen was repeated every four weeks. Patients were evaluated at the beginning of each course for objective response based on change in size of bidimensionally measurable solid lesions. Ten patients were removed from study due to progressive disease [defined as a 25% or greater increase in the size of measurable lesions]. This is an indication that the patients are doing quite poorly. Eight patients were removed from study due to stable disease after three courses (an indication that the disease is still not in remission). Six patients died during study course. This study is fairly typical of small scale chemotherapy trials conducted at the center.

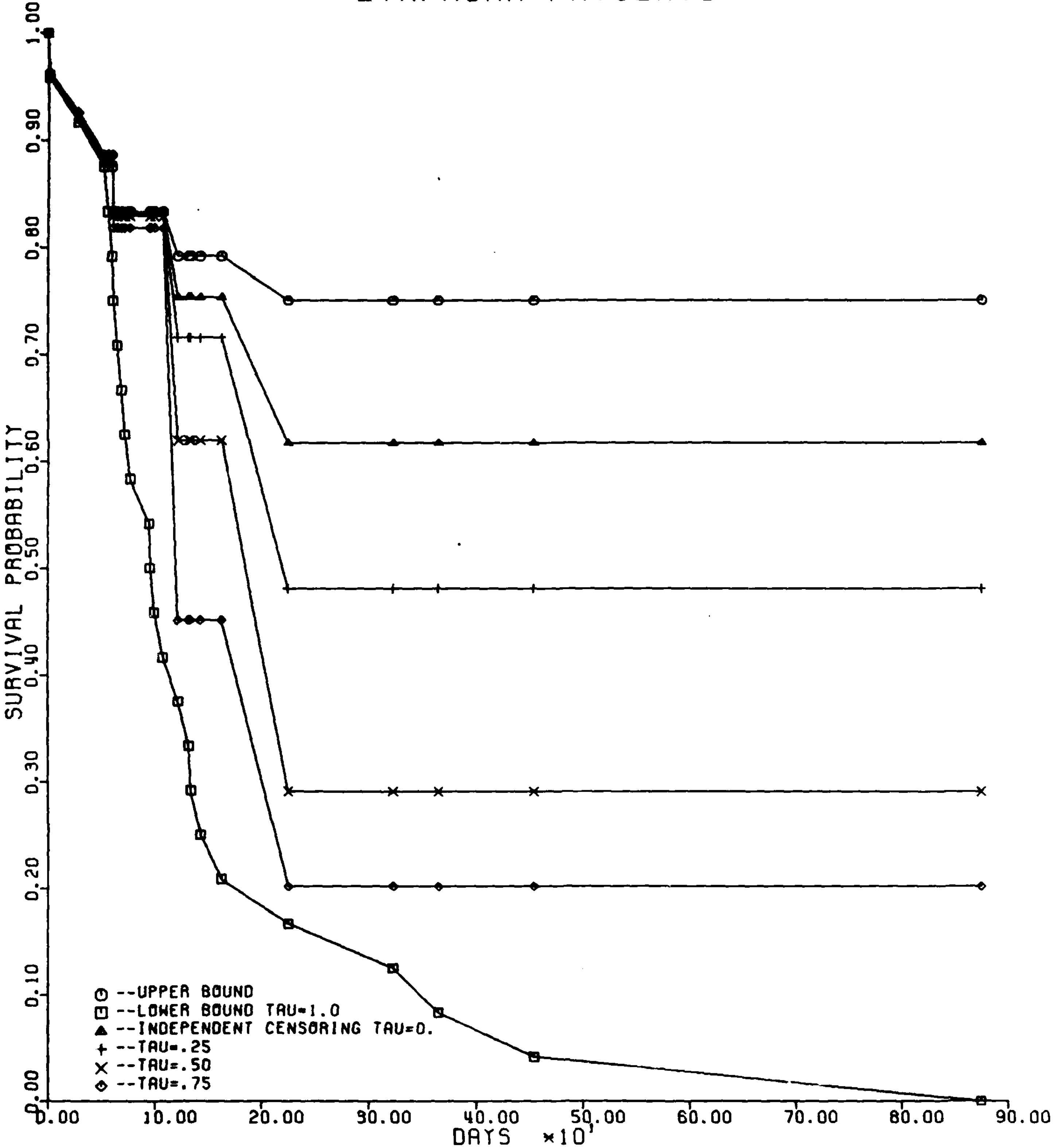
As in the previous example, here the times for patient removal from the study (censoring times) are clearly not independent of death times. The determination of bounds for τ, in this instance, is not as obvious. However, we think there are some reasonable possibilities which utilize the clinician's subjective understanding of the history and progression of the disease.

First, the clinician may be asked to select an upper and lower bound for τ from a set of classifications for association such as no association, $\tau = 0$; weak association, $\tau = .25$; moderate association, $\tau = .50$; strong association, $\tau = .75$; or perfect association $\tau = 1.0$. Second, the clinician may be asked to give a range of τ on a scale of 0 to 1. We note that this may be hard to obtain due to the physician's unfamiliarity with the concordance coefficient (see Kadane (1980) or Winkler (1980) for a discussion of this problem in the elicitation of opinion in the linear model framework). Third, if the clinician cannot make either of the first two judgments then he/she could be presented with a set of partial death or censoring times, such as (, 120), (77,), (82,), (, 113), etc. and asked to guess upper and lower bounds for each missing time, or as in Kadane et. al. (1980) to provide quantiles of the predictive distribution of the missing times. Based on this information an upper and lower τ could be computed. The program TROLL, described in Kadane et. al. (1980), which was designed to elicit opinion about the prior distributions of use in linear regression could be used to obtain a prior confidence interval for the value of the correlation coefficient ρ under an assumption of normality and then using the transformation $\tau=(2/\pi)\arcsin\sqrt{\rho}$ a range of values for τ is obtained. In an attempt to provide a more stable estimate of this range, a group of clinicians may be asked to make such estimations and the group's values may be used to obtain a range on τ.

Finally, charts of patients treated with this regimen might be reviewed and information from these charts be used to estimate the missing times. Information in the literature may also be used to estimate the missing times. Confidence interval methodology for τ may than be used to obtain a range of τ.

In Figure 2, we have plotted $\hat{S}(t)$ for various values of τ which correspond to a broad subjective range of τ. Based on discussion with physicians in the OSU Comprehensive Cancer Center, a subjective range for τ was from weak to strong association between times to progression or stable times and death times. This suggests using bounds corresponding to $\tau = .25$ to $\tau = .75$.

FIGURE 2
BOUNDS ON SURVIVAL FOR NON-HODGKINS LYMPHOMA PATIENTS



Acknowledgment

This work was supported by the Air Force Office of Scientific Research under contract AFOSR-82-0307 and NIH Research Grant Number P30CA 16058-14 from the National Cancer Institute.

References

Ahmed, A.H., Langberg, N.A., Leon, R.V., and Proschan, F. (1979). Partial Ordering of Positive Quadrant Dependence, with Applications. Florida State University Technical Report.

Basu, A.P., and Klein, J.P. (1982). Some Recent Results in Competing Risks Theory. Survival Analysis, Crowley and Johnson, eds. Institute of Mathematical Statistics, Hayward, CA, 216-229.

Clayton, D.G. (1978). A Model for Association in Bivariate Lifetables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence, Biometrika 65:141-151.

Cox, D.R. (1972). Regression Models and Life Table. Journal of the Royal Statistical Society B. 34:187-220.

David, H.A. and Moeschberger, M.L. (1978). Theory of Competing Risks. Griffin, London.

Fréchet, M. (1951). Surles Tableaux de Correlation Dont les Marges Sont Donnees. Annales de l'Universite' de Lyon, Section A, Series 3, 14:53-77.

Holford, T.R. (1980). The Analysis of Rates and of Survivorship Using Log-Linear Models. Biometrics 36:299-306.

Hutchinson, T.P. (1981). Compound gamma bivariate distributions. Metrika 28:263-271.

Kadane, J.B. (1980). Predictive and Structural Methods for Eliciting Prior Distributions. Studies in Bayesian Econometrics and Statistics in Honor of Harold Jeffreys, ed. A. Zellner, Amsterdam, North-Holland Publishing Co.

Kadane, J.B., Dickey, Jim, Winkler, R.L., Smith, W.S., and Peters, S.C. (1980). Interactive elicitation of opinion for a normal linear model. Journal of the American Statistical Association 75:845-854.

Kaplan, E.L. and Meier, P. (1958). Nonparametric Estimation from Incomplete Samples. JASA 53:457-481.

Kendall, M.G. (1962). Rank Correlation Methods, 3rd ed., Griffin, London.

Lagakos, S.W. (1979). General right censoring and its impact on the analysis of survival data. Biometrics 35:139-156.

Lagakos, S.W. and Williams, J.S. (1978). Models for Censored Survival Analysis: A cone class of variable - sum models. Biometrika 65:181-186.

Lindley, D.V. and Singpurwalla, N.D. (1984). Multivariate Distributions for the Reliability of a System of Components Sharing a Common Environment. George Washington University Technical Report GWU/IRRA/TR-841/.

Moeschberger, M.L. and Klein, J.P. (1984). Consequences of Departures From Independence in Exponential Series Systems. Technometrics, 26:277-284.

Nelson, W. (1972). Theory and Application of Hazard Plotting for Censored Failure Data. Technometrics 14:945-966.

Oakes, D. (1982). A Model for Association in Bivariate Survival Data. Journal of the Royal Statistical Society 44:412-422.

Peterson, A.V. (1976). Bounds for a Joint Distribution Function with Fixed Subdistribution Functions: Applications to Competing Risks, Proceedings of the National Academy of Sciences 73, 11-13.

Slud, E.V. and Rubenstein, L.V. (1983). Dependent Competing Risks and Summary Survival Curves. Biometrika 70:643-650.

Wei, L.J. and Gail, M.H. (1983). Nonparametric Estimation for a Scale-Change with Censored Observations. Journal of the American Statistical Association, 78:382-388.

Williams, J.S. and Lagakos, S.W. (1977). Models for Censored Survival Analysis: Constant Sum and Variable Sums Models. Biometrika, 64:215-224.

Winkler, R.L. (1980). Prior Information, Predictive Distribution and Bayesian Model-Building. Studies in Bayesian Econometrics and Statistics in Honor of Harold Jeffreys, ed. A. Zellner, Amsterdam: North-Holland Publishing Co., 95-109.

Appendix

Theorem 1. Let (X,Y) have the form (2.1) with marginals $S(t)$, $R(t)$ respectively. Let $\theta \geq 1$ be known. Then on the set where $S(t) > 0$ we have $\hat{S}_\theta(t) \to S(t)$ a.s.

Proof of Theorem 1:

For $\theta = 1$, the result is well known. Suppose that $\theta > 1$. Note that $\hat{Q}_1(t) \to Q_1(t) = P(T \leq t, x-y)$ a.s. and $F(u) \to F(u)$ a.s. by the strong law of large numbers. Since $\hat{S}_\theta(t)$ is a continuous function of $\int_0^t \frac{d\hat{Q}_1(u)}{[\hat{F}(u)]^\theta}$ in the support of $\hat{F}(u)$, it suffices to show

$$\int_0^t \frac{d\hat{Q}_1(u)}{[\hat{F}(u)]^\theta} \to \int_0^t \frac{dQ_1(u)}{[F(u)]^\theta} \quad \text{a.s.}$$

Now, after an integration by parts,

$$\int_0^t \frac{d\hat{Q}_1(u)}{[\hat{F}(u)]^\theta} = \frac{\hat{Q}_1(t)}{[\hat{F}(t)]^\theta} - \int_0^t \hat{Q}_1(u)\,d\left(\frac{1}{\hat{F}^\theta(u)}\right)$$

$$= \frac{\hat{Q}_1(t)}{[\hat{F}(t)]^\theta} - \int_0^t [\hat{Q}_1(u) - Q_1(u)]\,d\left(\frac{1}{\hat{F}^\theta(u)}\right) + \int_0^t Q_1(u)\,d\left(\frac{1}{\hat{F}^\theta(u)}\right)$$

$$= \frac{\hat{Q}_1(t) - Q_1(t)}{[\hat{F}(u)]^\theta} - \int_0^t [\hat{Q}_1(u) - Q_1(u)]\,d\left(\frac{1}{\hat{F}^\theta(u)}\right)$$

$$+ \int_0^t \frac{dQ_1(u)}{\hat{F}(u)^\theta}\ .$$

By the dominated convergence theorem

$$\lim_{n\to\infty} \int_0^t \frac{dQ_1(u)}{[\hat{F}(u)]^\theta} = \int_0^t \frac{dQ_1(u)}{[F(u)]^\theta} \quad \text{a.s.,}$$

$$\lim_{n\to\infty} \frac{\hat{Q}_1(t) - \hat{Q}_1(t)}{[\hat{F}\ (u)]^{\theta}} = 0 \quad \text{a.s.},$$

and

$$\lim_{n\to\infty} \sup \{|\hat{Q}_1(u) - Q_1(u)|\} = 0, \text{ a.s.}$$

Hence, applying the above results to (3.7), the result now follows: //

APPENDIX F

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY NA	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE NA	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION The Ohio State University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NM
6c. ADDRESS (City, State, and ZIP Code) 1314 Kinnear Road Columbus, Ohio 43212	7b. ADDRESS (City, State, and ZIP Code) AFOSR/NM Bldg 410 Bolling AFB DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-82-0307
8c. ADDRESS (City, State, and ZIP Code) AFOSR/NM Bldg 410 Bolling AFB DC 20332-6448	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO.: 61102F PROJECT NO.: 2304 TASK NO.: WORK UNIT ACCESSION NO.:

11. TITLE (Include Security Classification) A Test for Independence Based on Data From a Bivariate Series System

12. PERSONAL AUTHOR(S)
John P. Klein

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 9-1-82 TO 12-31-87	May 31, 1988	10

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Kendall's tau; testing for independent competing risks; series systems
XXXX	XXXXXXXXXX	XXX	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The problem of testing for independence of the component lifetimes when the components are linked in series is considered. To avoid the problem of nonidentifiability, the marginal component lifetimes are assumed to be known. In this setting, a modified version of Kendall's tau is proposed. This test statistic is obtained by replacing those component lifetimes which cannot be observed, due to system failure, by conditional probabilities computed under independence. A small scale simulation study of the power of this test shows the test has reasonable power for relatively small sample sizes.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Maj. Brian Woodruff	22b. TELEPHONE (Include Area Code) (202) 767-5027 22c. OFFICE SYMBOL NM

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

RELIABILITY AND QUALITY CONTROL
A.P. Basu (Editor)

A TEST FOR INDEPENDENCE BASED ON DATA FROM A BIVARIATE SERIES SYSTEM

John P. Klein

Department of Statistics
The Ohio State University
Columbus, Ohio
USA

The problem of testing for independence of the component lifetimes when the components are linked in series is considered. To avoid the problem of nonidentifiability, the marginal component lifetimes are assumed to be known. In this setting, a modified version of Kendall's tau is proposed. This test statistic is obtained by replacing those component lifetimes which cannot be observed, due to system failure, by conditional probabilities computed under independence. A small scale simulation study of the power of this test shows the test has reasonable power for relatively small sample sizes.

1. DEPENDENT SYSTEMS

A common assumption made in modeling series systems is that the component lifetimes are statistically independent. This assumption is also routinely made in analyzing data collected from series systems. Recently, Klein and Moeschberger (1983) and Moeschberger and Klein (1984) have shown that one may be appreciably misled by this independence assumption for certain bivariate exponential systems.

To illustrate the effects of this independence assumption, consider the following two models for the joint survival function of the component lifetimes (X,Y). The first model, due to Oakes (1982) has joint survival function

$$\overline{H}(x,y) = P(X>x,\ Y>y) = [\{1/\overline{F}(x)\}^{\theta-1} + \{1/\overline{G}(y)\}^{\theta-1} - 1]^{-1/(\theta-1)},\ \theta > 1 \tag{1.1}$$

where $\overline{F}(\cdot)$, $\overline{G}(\cdot)$ are the marginal survival functions of X and Y respectively. This distribution has a coefficient of concordance $\tau = (\theta-1)/(\theta+1)$ and $\theta = 1$ corresponds to independent component failure times. If $\lambda(x|Y=y)$ and $\lambda(x|Y>y)$ denote the conditional hazard functions for the conditional distributions of X given Y = y and given Y > y, respectively, then $\lambda(x|Y>y) = \theta\lambda(x|Y>y)$.

A second model, due to Gumbel (1960), has joint survival function

$$\overline{H}(x,y) = \overline{F}(x)\overline{G}(y)[1+\alpha(1-\overline{F}(x))(1-\overline{G}(y))],\ -1 < \alpha < 1 \tag{1.2}$$

This model has coefficient of concordance $\tau = 2\alpha/9$ which, unlike the Oakes model, may be both positive and negative.

To illustrate the importance of the independence assumption in modeling the system life, consider figures 1 and 2 where the 95th and 99th percentile of system life is plotted for the two models with exponential marginals. Here in all cases the first component has unit mean life. For the Gumbel model, the true percentile ranges from 80% to 115% of the percentile computed under independence, while in the Oakes model, the true percentile can be as much as twice as big as the percentile computed under independence when $\lambda_2 = \lambda_1$ and as much as 1.5 times as big when $\lambda_2 = 2$.

Since one may be appreciably misled by erroneously assuming independent component lifetimes, it is desirable to test this hypothesis based on data from series systems. However, if no assumptions about the underlying distribution of the component lifetimes is made, such a test is impossible due to the identifiability problem (see, e.g., Tsiatis (1978), Miller (1977), Basu and Klein (1982)). This is, given any set of observable information (such as system life, crude system life, etc.) collected from a series system with dependent component lifetimes, there exists a series system with independent component lifetimes with the same observable information (see Langberg, Proschan and Quinzi (1981)). However, this comparable system of independent random variables need not have the same marginal component life distributions as the dependent structure. In particular, the marginal distributions of the two systems are the same only for the class of constant sum models defined by Williams and Lagakos (1977).

In the next section, a modification of Kendall's (1938) test for independence is proposed. This test assumes that the marginal component life distributions are completely specified. This information could be obtained by testing each component separately, as is often done in the development stages of system design (see, e.g., Easterling and Prairie (1971), Mastran (1976), or Miyamura (1982)). In section 3, a simulation study compares the power of this test to the parametric tests based on the Oakes and Gumbel models.

2. THE TEST PROCEDURE

Suppose that n two component series systems are put on test. Let X_i, Y_i denote the potential (unobservable) failure times of the first and second components of the ith systems. We are not allowed to observe (X_i, Y_i) directly, but instead we observe $T_i = \min(X_i, Y_i)$, the system failure time and

$$I_i = \begin{cases} 1 \text{ if } T_1 = X_i, \\ 0 \text{ if } T_i = Y_i \end{cases} \quad \text{the cause of the system failure}$$

Also suppose that the maginal survival functions of X_i and Y_i, $\bar{F}(x) = P(X_i > x)$ and $\bar{G}(y) = P(Y_i > y)$, $i = 1, \ldots, n$ are known.

If we could observe both X_i and Y_i, then a test of independence, due to Kendall (1938), is to count the number of concordant pairs and the number of discoardant pairs. A pair (X_i, Y_i), (X_j, Y_j) is concordant if $X_i - X_j$ and $Y_i - Y_j$ have the same sign and is discordant if these differences have different signs. The test statistic is then the number of concordant pairs minus the number of discordant pairs.

If the data comes from a series system, then only T_i, I_i is observed. Suppose we consider a pair (T_i, I_i), (T_j, I_j) with $T_i < T_j$. If $I_i = 1$ and $I_j = 1$, then we know that $X_i = T_i < X_j = T_j$, and $X_i < Y_i$, $X_j < Y_j$. This pair would be concordant, regardless of the value of Y_j, if $T_i < Y_i < T_j$. If $Y_i > T_j$ concordance or discordance depends on the value of Y_j. Under the null hypothesis of independence, the conditional probablity that the pair is concordant is $[\bar{G}(T_i) - \bar{G}(T_j)]/\bar{G}(T_i) = P(T_i < Y < T_j | Y > T_i)$ since average concordance over the range $Y > T_j$ is 0. When $I_i = 1$ and $I_j = 0$, then $T_i = X_i < Y_j = T_j$, $X_i < Y_i$, $Y_j < X_j$. Here if $T_i < Y_i < T_j$, the pair would be concordant, and if $Y_i > T_j$, the pair would be discordant, whatever the value of X_j. Under independence, the conditional probablities of these two events are $[\bar{G}(T_i) - \bar{G}(T_j)]/\bar{G}(T_i)$ and $\bar{G}(T_j)/\bar{G}(T_i)$, respectively. Should $I_i = 0$, similar probabilities, involving F, could be obtained. This motivation suggests the following score function for $T_i < T_j$.

FIGURE 1
95TH PERCENTILE

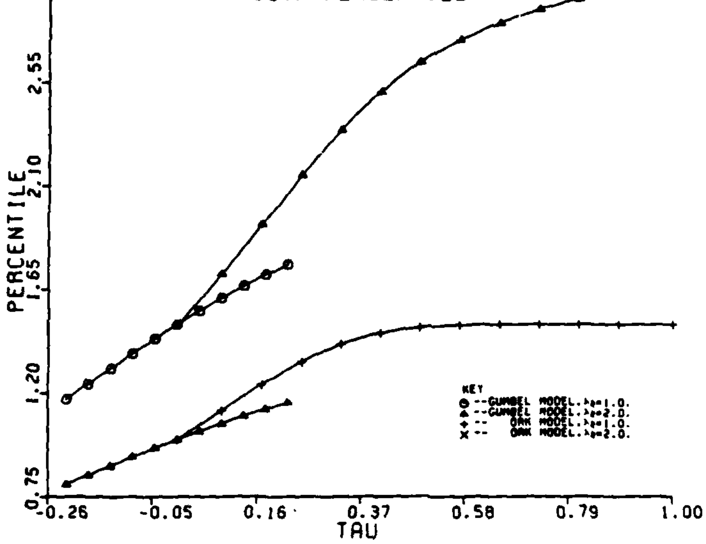
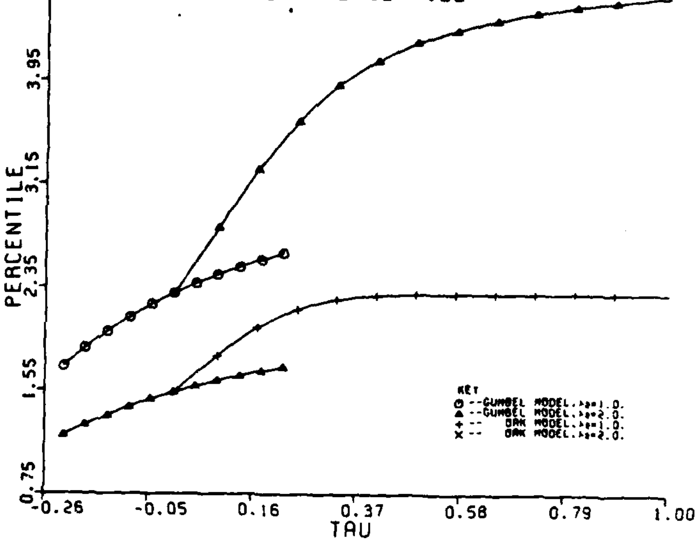


FIGURE 2
99TH PERCENTILE



$$\phi(T_i,I_i,T_j,I_j) = \begin{cases} [\bar{G}(T_i) - \bar{G}(T_j)]/\bar{G}(T_i) & \text{if } I_i = I_j = 1 \\ [\bar{F}(T_i) - \bar{F}(T_j)]/\bar{F}(T_i) & \text{if } I_i = I_j = 0 \\ [\bar{G}(T_i) - 2\bar{G}(T_j)]/\bar{G}(T_i) & \text{if } I_i = 1,\ I_j = 0 \\ [\bar{F}(T_i) - 2\bar{F}(T_j)]/\bar{F}(T_i) & \text{if } I_i = 0,\ I_j = 1 \end{cases} \quad (2.1)$$

and similarly for $T_i > T_j$.

The modified version of Kendall's test statistic is

$$\hat{\tau} = \sum_{1 \le i < j \le n} \sum \phi(T_i, I_i, T_j, I_j) / (n(n-1)/2) \quad (2.2)$$

To find the moments of τ, under independence, consider the pairs

(T_1, I_1), (T_2, I_2). Let $A_1 = \{T_1 < T_2, I_1 = I_2 = 1\}$, $A_2 = \{T_1 < T_2, I_1 = 1, I_2 = 0\}$, $A_3 = [T_1 < T_2, I_1 = 0, I_2 = 0\}$ and $A_4 = \{T_1 < T_2, I_1 = 0, I_2 = 1\}$. In terms of the unobservable component lifetimes, (X_i, Y_i), $A_1 = \{X_1 < X_2, X_2 < Y_2, X_1 < Y_1\}$ $A_2 = \{X_1 < Y_2, X_1 < Y_1, Y_2 < X_2\}$, $A_3 = \{Y_1 < Y_2, Y_1 < X_1, Y_2 < X_2\}$ and $A_4 = \{Y_1 < X_2, Y_1 < X_1, X_2 < Y_2\}$. Since, by symmetry, T_1 is equally likely to either smaller or larger than T_2, we have

$$(1/2)E(\phi(T_1,I_1,T_2,I_2)) = \int_{A_1} \frac{\bar{G}(x_1)-\bar{G}(x_2)}{\bar{G}(x_1)} dF(x_1)dF(x_2)dG(y_1)dG(y_2)$$
$$+ \int_{A_2} \frac{\bar{G}(x_1)-2\bar{G}(y_2)}{\bar{G}(x_1)} dF(x_1)dF(x_2)dG(y_1)dG(y_2)$$
$$+ \int_{A_3} \frac{\bar{F}(y_1)-\bar{F}(y_2)}{\bar{F}(y_1)} dF(x_1)dF(x_2)dG(y_1)dG(y_2) \quad (2.3)$$
$$+ \int_{A_4} \frac{\bar{F}(y_1)-2\bar{F}(x_2)}{\bar{F}(y_1)} dF(x_1)dF(x_2)dG(y_1)dG(y_2)$$
$$= J_1 + J_2 + J_3 + J_4 \text{ (say).}$$

Now, consider

$$J_1 + J_2 = \int_{-\infty}^{\infty} \{\int_x^{\infty} ([\bar{G}(x)-\bar{G}(y)]\bar{G}(y)dF(y) + \int_x^{\infty} [\bar{G}(x)-2\bar{G}(y)]\bar{F}(y)dG(y)\}dF(x) \quad (2.4)$$

Integrating the first inner integral in (2.4) by parts yields the negative of the second integral so that $J_1 + J_2 = 0$. Similar computations show that $J_3 + J_4 = 0$. Thus $E(\phi(T_1, I_1, T_2, I_2))$ and hence $E(\hat{\tau})$ are both 0. By similar computations, one can show that

$$n(n-1)V(\hat{\tau}) = (4/3) \int_{-\infty}^{\infty} \bar{G}(x)^2 dF(x) + (4/3) \int_{-\infty}^{\infty} \bar{F}(x)^2 dG(x)$$

$$- 4 \int_{-\infty}^{\infty} \bar{G}(x)^{-1} \int_{x}^{\infty} F(y)\bar{G}(y)^2 dG(y)dF(x)$$

$$- 4 \int_{-\infty}^{\infty} \bar{F}(x)^{-1} \int_{x}^{\infty} G(y)\bar{F}(y)^2 dF(y)dG(x)$$

$$+ 4(n-2)\{2/3 - 2\int_{-\infty}^{\infty} F(x)G(x)d(F(x) + G(x)) \tag{2.5}$$

$$- 2 \int_{-\infty}^{\infty} F(x)^2G(x)^2 d(F(x) + G(x))$$

$$+ 3 \int_{-\infty}^{\infty} F(x)^2G(x)dF(x) + 3\int_{-\infty}^{\infty} F(x)G(x)^2dG(x)\}, \text{ where } F(x) = 1 - \bar{F}(x) \text{ and } G(x) = 1 - \bar{G}(x).$$

The asymptotic normality of τ follows by the results of Hoeffding (1948). Hence, a test of independence versus dependence rejects if $|\hat{\tau}/\sqrt{V(\hat{\tau})}|$ is greater than the appropriate percentage point of a standard normal random variable. A test of independence versus positive dependence rejects if $\hat{\tau}/\sqrt{V(\hat{\tau})}$ is too large.

The variance of $\hat{\tau}$ (2.5) can be expressed explicitly in several cases.

Case 1. $\bar{F}(x) = \bar{G}(x)$. In this case (2.5) reduces to

$$V(\hat{\tau}) = (4n+7)/(30n(n-1)). \tag{2.6}$$

Case 2. (Lehmann structure) $\bar{F}(x) = \bar{G}(x)^{\alpha}$. Here (2.5) reduces to

$$n(n-1)V(\hat{\tau}) = 8\alpha[35\alpha + n(9\alpha^2 + 2\alpha + 9)]/[3(3\alpha + 1)(3 + \alpha)(2\alpha + 3)(3\alpha + 2)]. \tag{2.7}$$

Case 3. (X, Y exponential), $\bar{F}(X) = e^{-\lambda x}$, $\bar{G}(y) = e^{-\theta y}$, then (2.5) reduces to

$$n(n-1)V(\hat{\tau}) = \frac{8\lambda\theta[35\lambda\theta + n(9\lambda^2 + 2\lambda\theta + 9\theta^2)]}{3(3\lambda + \theta)(\lambda + 3\theta)(2\lambda + 3\theta)(3\lambda + 2\theta)} \tag{2.8}$$

When the true values of $\bar{F}$, $\bar{G}$ are misspecified, then $E(\hat{\tau})$ is not zero. If the true component lifetime distributions are $\bar{F}$, $\bar{G}$ but $\bar{F}^{\alpha}$, $\bar{G}^{\beta}$ are used in formula (2.3), then one can show that, under independence,

$$E(\hat{\tau}) = 2(1-\beta)\int_{-\infty}^{\infty} \bar{G}(x)^{1-\beta} \int_{x}^{\infty} \bar{F}(y)\bar{G}(y)^{\beta} dG(y)dF(x)$$

$$+ 2(1-\alpha) \int_{-\infty}^{\infty} \bar{F}(x)^{1-\alpha} \int_{x}^{\infty} \bar{G}(y)\bar{F}(y)^{\alpha} dF(y)dG(x), \ \alpha, \beta > 0.$$

If $\bar{F}(\cdot) = \bar{G}(\cdot)$, then $E(\hat{\tau}) = (\beta-1)/2(\beta+2) + (\alpha-1)/2(\alpha+2)$.

If $\bar{F}(x) = \bar{G}(x)^{\theta}$, then $E(\hat{\tau}) = (\theta/(\theta+1)) \{(\alpha-1)/(\theta+\alpha\theta+1) + (\beta-1)/(\theta+\beta+1)\}$.

Similar expressions can be obtained for the null variance of $\hat{\tau}$.

3. SIMULATION STUDY

To study the effectiveness of the modified Kendall's τ described in section 2, a simulation study was conducted. The study was performed by generating 1000 samples of $n = 20$ or 40 series systems with exponentially distributed component life times, $\bar{F}(x) = e^{-x}$ and $\bar{G}(y) = e^{-\lambda_2 y}$, $\lambda_2 = 1, 2$. Both the Oakes joint distribution (1.1) and the bivariate Gumbel distribution (1.2) were used. The bivariate observations from the Oakes distribution were generated using the technique described in section 2 of that paper. To generate Gumbel random variables with marginal survival functions $\bar{F}(x)$, $\bar{G}(y)$, let U_1, U_2 be independent uniform random deviates. Note that

$$\bar{F}(x|y) = P(X > x|Y = y) = \bar{F}(x)(1+\alpha-2\alpha\bar{G}(y)) - \alpha\bar{F}(x)^2(1-2\bar{G}(y)). \tag{3.1}$$

Let $U_1 = \bar{G}(y)$ and $U_2 = \bar{F}(x|y) = \bar{F}(x)[1+\alpha-2\alpha U_1] - \alpha\bar{F}(x)^2(1-2U_1)$.

Solving this equation for $\bar{F}(x)$ yields

$$\bar{F}(x) = U^* = [(1+\alpha(1-2U_1)) - \alpha(1+\alpha^2(1-2U_1)^2 + 2\alpha(1-2U_1)(1-2U_1))^{1/2}] \Big/ 2\alpha(1-2U_1),\ U_1 \neq 1/2 \tag{3.2}$$

which is the root which lies in the interval [0,1]. If $U_1 = 1/2$, then $U^* = U_2$. The pair (X, Y) is then found by $X = \bar{F}^{-1}(U^*)$, $Y = \bar{G}^{-1}(U_1)$.

For the purpose of comparison, the parametric tests for independence, based on the efficient scores statistics, for the Gumbel and Oakes model were obtained. Consider first the Gumbel model (1.2). Using the notation in section 2, the observable crude density for $I = 1$ is

$$-(d/dt)\, P(T>t, I=1) = q_1(t) = f(t)\bar{G}(t)[1+\alpha(1-\bar{F}(t)-2\bar{G}(t) + \bar{F}(t)\bar{G}(t))]$$

where $f(t) = -(d/dt)\,\bar{F}(t)$, and a similar expression for $q_0(t)$ when $I = 0$. Based on a sample of n series systems, the likelihood function is

$$L(\alpha) = \prod_{j=1}^{n} q_1(t_j)^{I_j}\, q_0(t_j)^{1-I_j}. \tag{3.3}$$

After some simplification, the likelihood equation at $\alpha = 0$ is

$$(d/d\alpha)\ \ln L(\alpha)\Big|_{\alpha=0} = W = n - \sum_{j=1}^{n} \bar{F}(t_j) + \bar{G}(t_j) - \bar{F}(t_j)\bar{G}(t_j) + I_j\bar{G}(t_j) + (1-I_j)\bar{F}(t_j)$$

Also, when X and Y are exponential with parameters λ_1, λ_2, respectively, then

$\sigma^2{}_G = -E(d^2 \ln(\alpha)/d\alpha^2) = n\ \{7/3 + (\lambda_1 + 4\lambda_2)/(\lambda_1 + 3\lambda_2)$
$+ (\lambda_2 + 4\lambda_1)/(\lambda_1 + 3\lambda_2) - 4[(\lambda_1 + 2\lambda_2)/(2\lambda_1 + 3\lambda_2) + (2\lambda_1 + \lambda_2)/$
$(3\lambda_1 + 2\lambda_2)]\}$.

The efficient scores test of the hypothesis $\alpha = 0$ is then W/σ_G which has an approximate standard normal distribution when $\alpha = 0$.

For the Oakes model (1.1), the crude density functions are of the form

$$q_1(t) = f(t)/\bar{F}^{\theta}(t)\ \bar{S}(t)^{\theta} \text{ where} \tag{3.4}$$

$$\bar{S}(t)^{\theta} = [\{1/\bar{F}(t)\}^{\theta-1} + \{1/\bar{G}(t)\}^{\theta-1} - 1]^{-1/(\theta-1)} \tag{3.5}$$

From (3.3), the log likelihood for θ is

$$\ln L(\theta) = -\sum_{j=1}^{n} \theta\ I_j a_j + (1-I_j)b_j + \theta\ (\ln D_j)/(\theta-1) \tag{3.6}$$

where $a_j = -\ln\bar{F}(t_j)$, $b_j = -\ln\bar{G}(t_j)$, and $D_j = \{e^{(\theta-1)a_j} + e^{(\theta-1)b_j} - 1\}$.

The likelihood equation is

$$(d/d\theta)\ln L(\theta) = \sum_{j=1}^{n} I_j a_j + (1-I_j)b_j - (1 + 1/(\theta-1))\sum_{j=1}^{n} [a_j\ e^{(\theta-1)a_j} + b_j\ e^{(\theta-1)b_j}]/D_j$$

$$+ (1/(\theta-1)^2) \sum_{j=1}^{n} \ln D_j \tag{3.7}$$

$$\text{which is equal to } V = \sum_{j=1}^{n} (a_j I_j + (1-I_j)b_j) + \sum_{j=1}^{n} (a_j b_j - a_j - b_j) \tag{3.8}$$

as $\theta \to 1^+$. The $E(-(d^2/d\theta^2)\ \ln L(\theta))$ at $\theta = 1$, under the exponential model is

$\sigma^2_0 = 2n\ \lambda_1\ \lambda_2/(\lambda_1 + \lambda_2)^2$ so the resulting score statistic is

$$Z = (\sum_{j=1}^{n} I_j(\lambda_1 T_j) + (1-I_j)\ \lambda_2 T_j + \sum_{j=1}^{n} \lambda_1\ \lambda_2 T^2_j - (\lambda_1 + \lambda_2)T_j)) \cdot (\lambda_1 + \lambda_2)/(2n\ \lambda_1\ \lambda_2)^{1/2}$$

which is approximately standard normal for large n when $\theta = 1$.

The results of this study are reported in table 1. From this table it seems like the modified τ test has reasonably good power when compared to the parametric tests, although comparison with the Oakes score test is hard since the significance level of that test is inflated. Also the test based on the Gumbel scores has comparable power when the data is from the Oakes model. A test for normality done on the samples where the components were independent accepted the normality assumption for the modified τ test.

Table 2 reports the observed number of rejections when the component parameters are estimated based on independent samples of size 50 for each component. A .05 significance level was used. Here, when $\lambda_1 = \lambda_2$, all tests have inflated levels. When $\lambda_1 \neq \lambda_2$ the tests are conservative. All tests have comparable power when $\lambda_1 = \lambda_2$, however, the modified τ test has significantly higher power when $\lambda_1 \neq \lambda_2$.

In addition to the power of our modified test, the $E(\hat{\tau})$ was estimated for each sample. Except in the independence case, the simulation showed that $E(\hat{\tau}) = .35\tau$, suggesting τ is of limited use as a point estimator of τ.

TABLE 1
Estimated Power Using True Parameter Values
Based On 1000 Replications

MODEL		n	τ	MODIFIED τ α = .05	MODIFIED τ α = .025	OAKES SCORE α = .05	OAKES SCORE α = .025	GUMBEL SCORE α = .05	GUMBEL SCORE α = .025
Independent	(λ_2=1)	20	0	50	25	71+	52+	58	37
Independent	(λ_2=1)	40	0	42	21	57	34	54	36
Independent	(λ_2=2)	20	0	53	25	74+	52+	62	34
Independent	(λ_2=2)	40	0	55	32	74+	49+	69	37
Gumbel	(λ_2=1)	20	.125	99	55	115	88	133	78
Gumbel	(λ_2=1)	40	.125	158	96	159	124	192	124
Gumbel	(λ_2=2)	20	.125	119	74	128	88	141	100
Gumbel	(λ_2=2)	40	.125	158	88	146	111	176	116
Gumbel	(λ_2=1)	20	.222	182	117	172	130	245	175
Gumbel	(λ_2=1)	40	.222	283	199	239	179	323	257
Gumbel	(λ_2=2)	20	.222	188	110	160	130	205	143
Gumbel	(λ_2=2)	40	.222	278	181	221	159	316	237
Oakes	(λ_2=1)	20	.125	170	114	236	202	184	137
Oakes	(λ_2=1)	40	.125	224	154	327	273	247	185
Oakes	(λ_2=2)	20	.125	166	101	231	207	179	125
Oakes	(λ_2=2)	40	.125	228	148	313	253	248	166
Oakes	(λ_2=1)	20	.25	318	243	421	377	379	295
Oakes	(λ_2=1)	40	.25	484	394	614	551	510	443
Oakes	(λ_2=2)	20	.25	334	223	386	335	354	273
Oakes	(λ_2=2)	40	.25	513	338	555	483	522	407
Oakes	(λ_2=1)	20	.50	638	535	704	670	680	606
Oakes	(λ_2=1)	40	.50	880	802	903	875	881	851
Oakes	(λ_2=2)	20	.50	657	589	615	547	674	593
Oakes	(λ_2=2)	40	.50	894	823	816	772	873	820
Oakes	(λ_2=1)	20	.75	799	722	803	763	858	795
Oakes	(λ_2=1)	40	.75	973	946	983	968	925	900
Oakes	(λ_2=2)	20	.75	899	847	699	631	823	763
Oakes	(λ_2=2)	40	.75	995	989	924	884	985	961

TABLE 2
Estimated Power Using Estimated Parameter Values
and 0.05 Significance Level

MODEL	n	τ	Modified τ $\lambda_1=1$	Modified τ $\lambda_2=2$	Oakes Score $\lambda_1=1$	Oakes Score $\lambda_2=2$	Gumbel Score $\lambda_1=1$	Gumbel Score $\lambda_2=2$
Independent	20	0	83	18	96	1	89	1
Independent	40	0	64	17	83	1	81	1
Gumbel	20	.125	136	42	141	3	159	4
Gumbel	40	.125	214	30	200	3	242	4
Gumbel	20	.22	209	68	201	6	255	10
Gumbel	40	.22	331	62	276	10	360	9
Oakes	20	.125	202	48	263	40	211	8
Oakes	40	.125	276	54	354	13	284	1
Oakes	20	.25	327	146	430	55	388	29
Oakes	40	.25	513	156	628	48	542	20
Oakes	20	.50	638	400	699	93	655	92
Oakes	40	.50	858	558	827	99	865	102
Oakes	20	.75	781	737	793	76	828	84
Oakes	40	.75	956	916	947	112	962	138

ACKNOWLEDGEMENTS

The author would like to thank Professor M. L. Moeschberger of The Ohio State University and P. Laud of Northern Illinois University for their constructive comments. This research was supported by contract AFOSR-82-0307 for the Air Force Office of Scientific Research.

REFERENCES

[1] Basu, A. P. and Klein, J. P. (1982). "Some Recent Results on Competing Risks Theory," in Survival Analysis, eds., J. Crowley and R. Johnson, Hayward, CA.: Institute of Mathematical Statistics, 216-229.

[2] David, H. A. and Moeschberger, M. L. (1978). The Theory of Competing Risks, New York: MacMillan Publishing Co., Inc.

[3] Easterling, R. G. and Prairie, R. R. (1971). "Combining Component and System Information," Technometrics, 13, 271-280.

[4] Gumbel, E. J. (1960). "Bivariate Exponential Distributions," Journal of the American Statistical Association, 55, 698-707.

[5] Hoeffding, W. (1948). "A Class of Statistics with Asymptotically Normal Distributions," Annals of Mathematical Statistics, 19, 293-325.

[6] Kendall, M (1938). "A New Measure of Rank Correlation," Biometrika, 30, 81-93.

[7] Klein, J. P. and Moeschberger, M. L. (1983). "Consequences of Assuming Independence in a Bivariate Exponential Series System," Ohio State University Technical Report.

[8] Langberg, N., Proschan, F., and Quinzi, A. J. (1978). "Estimating Dependent Life Lengths, with Applications to the Theory of Competing Risks," Annals of Statistics, 9, 157-167.

[9] Mastran, D. V. (1976). "Incorporating Component and System Data into the Same Assessment: A Bayesian Approach," Operation Research Society Journal, 24, 491-499.

[10] Miller, D. R. (1977). "A Note on Independence of Multivariate Lifetimes in Competing Risks," Annals of Statistics, 5, 576-579.

[11] Miyamura, T. (1982). "Estimating Component Failure Rates from Combined Component and System Data: Exponentially Distributed Component Lifetimes," Technometrics, 24, 313-318.

[12] Moeschberger, M. L. and Klein, J. P. (1984). "Robustness to Departures from Independence in Exponential Series Systems," Technometrics, 26, 277-283.

[13] Oakes, D. (1982). "A Model for Association in Bivariate Survival Data," Journal of the Royal Statistical Society B, 44, 414-442.

[14] Tsiatis, A. (1975). "A Nonidentifiability Aspect of the Problem of Competing Risks," Proceedings of the National Academy of Sciences, 72, 20-22.

[15] Williams, J. S. and Lagakos, S. W. (1977). "Models for Censored Survival Analysis, Constant Sum and Variable Sum Models," Biometrika, 64, 215-224.

APPENDIX G

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY NA	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE NA	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION The Ohio State University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NM
6c. ADDRESS (City, State, and ZIP Code) 1314 Kinnear Road Columbus, Ohio 43212	7b. ADDRESS (City, State, and ZIP Code) AFOSR/NM Bldg 410 Bolling AFB DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-82-0307
8c. ADDRESS (City, State, and ZIP Code) AFOSR/NM Bldg 410 Bolling AFB DC 20332-6448	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO.: 61102F PROJECT NO.: 2304 TASK NO.: WORK UNIT ACCESSION NO.:

11. TITLE (Include Security Classification) A Comparison of Several Methods of Estimating the Survival Function when There Is Extreme Right Censoring.

12. PERSONAL AUTHOR(S)
M.L. Moeschberger and John P. Klein

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 9-1-82 TO 12-31-87	May 31, 1988	7

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP
XXXX	XXXXXXXXXX	XXX

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Adjusted Kaplan-Meier survival estimation; bias of survival function; life-testing; right censoring; survival analysis

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

When there is extreme censoring on the right, the Kaplan-Meier product-limit estimator is known to be a biased estimator of the survival function. Several modifications of the Kaplan-Meier estimator are examined and compared with respect to bias and mean squared error.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Maj. Brian Woodruff	22b. TELEPHONE (Include Area Code) (202) 767-5027 22c. OFFICE SYMBOL NM

BIOMETRICS 41, 253–259
March 1985

A Comparison of Several Methods of Estimating the Survival Function when There Is Extreme Right Censoring

M. L. Moeschberger[1] and John P. Klein[2]

Departments of Preventive Medicine[1] and Statistics[2], The Ohio State University,
320 West 10th Avenue, Columbus, Ohio 43210, U.S.A.

SUMMARY

When there is extreme censoring on the right, the Kaplan–Meier product-limit estimator is known to be a biased estimator of the survival function. Several modifications of the Kaplan–Meier estimator are examined and compared with respect to bias and mean squared error.

1. Introduction

In human and animal survival studies, as well as in life-testing experiments in the physical sciences, one method of estimating the underlying survival distribution (or the reliability of a piece of equipment) which has received widespread attention is the Kaplan–Meier product-limit estimator (Kaplan and Meier, 1958).

For the situation in which the longest time an individual is in a study (or on test) is not a failure time, but rather a censored observation, it is well known that there are many complex problems associated with any statistical analysis (Lagakos, 1979). In particular, the Kaplan–Meier product-limit estimator is biased on the low side (Gross and Clark, 1975). In the case of many censored observations larger than the largest observed failure time, this bias tends to be worse. Estimated mean survival time and selected percentiles, as well as other quantities dependent on knowledge of the tail of the survival function, will also exhibit such biases.

A practical situation which motivates this study is a large-scale animal experiment conducted at the National Center for Toxicological Research (NCTR), in which mice were fed a particular dose of a carcinogen. The goal of the experiment was to assess the effects of the carcinogen on survival and on age-specific tumor incidence. Toward this end, mice were randomly divided into three groups and followed until death or until a prespecified group censoring time (280, 420, or 560 days) was reached, at which time all those still alive in a given group were sacrificed. Often there were many surviving mice in all three groups at the sacrifice times.

In general, we consider an experiment in which n individuals are under study and censoring is permitted. Let $t_{(1)}, \ldots, t_{(m)}$ denote the m ordered failure times of those m individuals whose failure times are actually observed ($t_{(1)} \leq \cdots \leq t_{(m)}$). The remaining $n - m$ individuals have been censored at various points in time. It will be useful to introduce the notation S_j to denote the number of survivors just prior to time $t_{(j)}$; that is, S_j is the number of individuals still under observation at time $t_{(j)}$, including the one that died at $t_{(j)}$. Then the Kaplan–Meier product-limit estimator (assuming no ties among the $t_{(j)}$) of

Key words: Adjusted Kaplan–Meier survival estimation; Bias of survival function; Life-testing; Right censoring; Survival analysis.

So, by the theorem on order statistics stated at the beginning of this section, the conditional distribution of $T_{(u)}$, given $T_{(n-n_c)} = t_{(n-n_c)}$ $(u = n - n_c + 1, \ldots, n)$ will be approximated by the $(u - n + n_c)$th order statistic in a sample of n_c drawn from (3). For simplicity, let $j = u - n + n_c$, so that $j = 1, \ldots, n_c$. Now the expected value of the jth order statistic from (3) is

$$\mathrm{E}(T_{j:n_c}) = n_c\binom{n_c - 1}{j - 1}\int_{t_c}^{\infty} t[P_T(t)]^{j-1}[\bar{P}_T(t)]^{n_c-j+1}(kt^{k-1}/\theta)\,dt$$

$$= n_c\binom{n_c - 1}{j - 1}\int_{0}^{\infty} (y^k + t_c^k)^{1/k}[P(y)]^{j-1}[\bar{P}(y)]^{n_c-j+1}(ky^{k-1}/\theta)\,dy \tag{4}$$

where $\bar{P}(y) = \exp(-y^k/\theta)$, $y = (t^k - t_c^k)^{1/k} \geq 0$ and $T_{j:n_c}$ is the jth order statistic in a sample of size n_c. Equation (4) can also be written as

$$\mathrm{E}(T_{j:n_c}) = n_c\binom{n_c - 1}{j - 1}\int_{0}^{\infty} (\theta z^k + t_c^k)^{1/k}[P(z)]^{j-1}[\bar{P}(z)]^{n_c-j+1}kz^{k-1}\,dz \tag{5}$$

where $\bar{P}(z) = \exp(-z^k)$, $z = (y/\theta)^{1/k} \geq 0$. Now $\mathrm{E}(T_{j:n_c})$ may be crudely estimated by

$$\{\hat{\theta}[\mathrm{E}(Z_{j:n_c})]^{\hat{k}} + t_c^{\hat{k}}\}^{1/\hat{k}} \tag{6}$$

where $\mathrm{E}(Z_{j:n_c})$ is the expected value of the jth order statistic from a sample of size n_c determined from Harter's (1969) tables or recurrence relation, and $\hat{\theta}$ and $\hat{k}$ are maximum likelihood estimators of θ and k, respectively.

These n_c estimated expected order statistics may then be treated as "observed" lifetimes in adjusting (or "completing") the estimated survival function computed in (1). The area under the estimated survival function up to t_c remains unchanged. The area under the extended estimated survival function based on the n_c estimated expected order statistics is then added to the initial area to obtain a more precise estimate of $\bar{P}(t)$ [estimated order statistic (EOS) extension].

2.2 *Weibull Maximum Likelihood Techniques*

A straightforward approach to completing $\hat{\bar{P}}(t)$ is to set

$$\hat{\bar{P}}(t) = \exp(-t^{\hat{k}}/\hat{\theta}) \qquad \text{for} \quad t > t_c. \tag{7}$$

Estimates of k and θ based on all observations can be obtained by either the maximum likelihood (WTAIL) or the least squares method. However, our study found the completion using maximum likelihood estimators was always better in terms of bias and mean squared error.

One suggestion for ostensibly improving this estimator would be to "tie" the estimated tail to the product-limit estimator at t_c. Two methods were attempted to accomplish this goal. First, the likelihood was maximized with respect to k and θ subject to the constraint that $\exp(-t_c^k/\theta) = \hat{\bar{P}}(t_c)$. This method will be referred to as the restricted MLE tail probability estimate (RWTAIL extension). Second, a scale-shift was performed on the tail probability in (7) to tie it to the product-limit estimator. This method led to higher biases and mean squared errors of the survival function and will be dropped from further discussion in this paper.

2.3 *BHK-Type Methods*

The Brown–Hollander–Korwar completion of the product-limit estimator sets

$$\hat{\bar{P}}(t) = \exp(-t/\theta^*) \qquad \text{for} \quad t > t_c \tag{8}$$

3. A Comparison of the Various Methods

A simulation study of data such as that collected at NCTR was performed. Three groups of 48 lifetimes were simulated with all testing stopping at 280, 420, and 560 days, respectively, for the three groups. Distributions with mean survival times of 400, 500, and 600 days were used. The generated lifetimes greater than or equal to the sacrifice time for each particular group were considered as censored. The remaining set of observed lifetimes, along with the number censored at the three sacrifice times, constituted a single sample. For each of the distributions studied, 1000 such samples were generated. Weibull distributions with shape parameters .5, decreasing failure rate, 1, constant failure rate, and 4,

Table 2

Bias/100 (and MSE/100^2) for estimating 90th percentile for various methods of completion

Distribution	μ	K–M	BHK extension	Estimated order statistic extension	Weibull WTAIL extension	Restricted Weibull RWTAIL extension
Weibull	400	−5.017[w]	−2.858	1.691	.234[b]	.458
		(25.185)[w]	(9.358)	(16.424)	(7.524)[b]	(10.812)
$k = .5$	500	−7.655[w]	−4.620	1.897	.418[b]	.642
		(58.604)[w]	(22.711)	(24.276)	(14.319)[b]	(21.442)
	600	−10.306[w]	−6.390	2.213	.734[b]	1.064
		(106.21)[w]	(42.449)	(36.895)	(25.419)[b]	(37.911)
	400	−3.610[w]	.064[b]	.248	.084	.067
		(13.035)[w]	(1.892)[b]	(2.423)	(1.980)	(2.945)
$k = 1$	500	−5.913[w]	.096[b]	.289	.121	.306
		(34.963)[w]	(2.995)[b]	(4.681)	(4.361)	(5.903)
	600	−8.216[w]	.244[b]	.610	.418	.550
		(67.459)[w]	(4.198)[b]	(9.247)	(8.331)	(10.792)
	400	−.045	.098[w]	−.007[b]	−.037	−.011
		(.038)[b]	(.236)[w]	(.060)	(.047)	(.063)
$k = 4$	500	−1.195	5.324[w]	−.031	−.026	.024[b]
		(1.429)	(33.091)[w]	(.146)	(.141)[b]	(.177)
	600	−2.554	17.913[w]	.120	.090	.068[b]
		(6.524)	(355.02)[w]	(.794)	(.676)	(.641)[b]
Lognormal	400	−2.628[w]	−.044[b]	−1.263	−1.758	−.967
		(6.908)[w]	(1.526)[b]	(1.979)	(3.407)	(1.673)
$k = 1$	500	−4.680[w]	.213[b]	−2.354	−2.718	−1.908
		(21.902)[w]	(2.708)[b]	(6.153)	(7.909)	(4.751)
	600	−6.736[w]	.759[b]	−3.507	−3.766	−2.980
		(45.373)[w]	(4.764)[b]	(13.123)	(14.981)	(10.257)
	400	−.085	.161	−.038	−.162[w]	−.024[b]
		(.060)[b]	(.409)[w]	(.081)	(.065)	(.093)
$k = 4$	500	−1.251	3.722[w]	−.584	−.657	−.484[b]
		(1.566)	(17.654)[w]	(.403)	(.495)	(.318)[b]
	600	−2.621	13.695[w]	−1.214	−1.236	−1.158[b]
		(6.872)	(210.30)[w]	(1.616)	(1.662)	(1.498)[b]
Bathtub	400	−3.629[w]	−.177	.053[b]	−.104	.105
		(13.167)[w]	(1.717)[b]	(2.052)	(2.058)	(3.190)
$p = .1$	500	−6.068[w]	−.457	−.071	−.208	.004[b]
		(36.826)[w]	(2.955)[b]	(4.702)	(3.619)	(5.245)
	600	−7.997[w]	−.318	.043	−.244	−.014[b]
		(63.954)[w]	(4.330)[b]	(7.786)	(7.608)	(9.923)
	400	−.347	.143[b]	.276	1.154[w]	.981
		(.273)[b]	(.844)	(1.078)	(3.877)	(4.747)[w]
$p = .4$	500	−1.425	.521[b]	.764	1.699	1.718[w]
		(2.035)	(1.540)[b]	(2.067)	(8.574)	(10.714)[w]
	600	−3.554[w]	−.137	.132[b]	2.304	2.450
		(12.628)[w]	(1.804)[b]	(2.352)	(17.530)	(22.456)

[b] Best estimation method.
[w] Worst estimation method.

Gross, A. J. and Clark, V. A. (1975). *Survival Distributions: Reliability Applications in the Biomedical Sciences*. New York: Wiley.

Harter, H. L. (1969). *Order Statistics and Their Use in Testing and Estimation. Vol.* 2. Aerospace Research Laboratories, Office of Aerospace Research, United States Air Force.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457–481.

Lagakos, S. W. (1979). General right censoring and its impact on the analysis of survival data. *Biometrics* **35**, 139–156.

White, J. S. (1969). The moments of log-Weibull order statistics. *Technometrics* **11**, 373–386.

Received June 1980; *revised May* 1984.

APPENDIX H

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY
NA

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE
NA

3. DISTRIBUTION/AVAILABILITY OF REPORT
Approved for Public Release; Distribution Unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S)

5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION
The Ohio State University

6b. OFFICE SYMBOL *(If applicable)*

7a. NAME OF MONITORING ORGANIZATION
AFOSR/NM

6c. ADDRESS *(City, State and ZIP Code)*
1314 Kinnear Road
Columbus, Ohio 43212

7b. ADDRESS *(City, State and ZIP Code)*
Bldg. 410
Bolling AFB, DC 20332-6448

8a. NAME OF FUNDING/SPONSORING ORGANIZATION
AFOSR

8b. OFFICE SYMBOL *(If applicable)*
NM

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
AFOSR-82-0307

8c. ADDRESS *(City, State and ZIP Code)*
Bldg. 410
Bolling AFB, DC

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
6.1102F	2304		

11. TITLE *(Include Security Classification)* A Partially Parametric Estimator of Survival in the Presence of Randomly Censored Data

12. PERSONAL AUTHOR(S)
John P. Klein, Shin-Chang Lee, and M. L. Moeschberger

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Final	FROM 9-1-82 TO 12-31-87	May 31, 1988	18

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB. GR.
XXXXXXXXXXXXXX		

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
Censored observations, Survival function estimation, Product-limit estimator.

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

Many biological or medical experiments have as their goal to estimate the survival function of a specified population of subjects when the time to the specified event may be censored due to loss to follow-up, the occurrence of another event which precludes the occurrence of the event of interest, or the study ends before the event of interest occurs. This paper suggests an improvement of the Kaplan-Meier product-limit estimator when the censoring mechnaism is random. The proposed estimator treats the uncensored observations nonparametrically and uses a parametric model only for the censored observations. One version of this proposed estimator always has a smaller bias and mean squared error than the product-limit estimator. An example estimating the survival function of patients enrolled in The Ohio State University Bone Marrow Transplant Program is presented.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL
Brian W. Woodruff, Major

22b. TELEPHONE NUMBER *(Include Area Code)*
(202) 767-5027

22c. OFFICE SYMBOL
AFOSR/NM

A Partially Parametric Estimator of Survival in the Presence of Randomly Censored Data

by

John P. Klein
Department of Statistics
The Ohio State University
Columbus, Ohio 43210

Shin-Chang Lee
Department of Mathematics and Computer Science
Florida International University
Miami, Florida

M. L. Moeschberger
Department of Preventive Medicine
The Ohio State University
Columbus, Ohio 43210

Summary

Many biological or medical experiments have as their goal to estimate the survival function of a specified population of subjects when the time to the specified event may be censored due to loss to follow-up, the occurrence of another event which precludes the occurrence of the event of interest, or the study being terminated before the event of interest occurs. This paper suggests an improvement of the Kaplan-Meier product-limit estimator when the censoring mechanism is random. The proposed estimator treats the uncensored observations nonparametrically and uses a parametric model only for the censored observations. One version of this proposed estimator always has a smaller bias and mean squared error than the product-limit estimator. An example estimating the survival function of patients enrolled in The Ohio State University Bone Marrow Transplant Program is presented.

Key words: Censored Observations, Survival Function Estimation, Product-limit Estimator

1. INTRODUCTION

The problem of analyzing data from experiments or clinical trials for which the time to occurrence of the primary event of interest (such as death, appearance of a tumor or other morbid condition, component failure, etc.) may not be directly observable due to the occurrence of some competing risk of removal from the study (such as loss to follow-up, early termination of the study, failure of a different cause or component than the one of interest, etc.) arises in biological, medical, or engineering sciences. In engineering applications, the data may be censored when one analyzes experiments involving series systems with several failure modes, testing of field equipment with a fixed test time and random or staggered entry, Type I or Type II censored life tests, or any combination thereof. In medical applications, data may be censored if we are conducting a clinical trial with fixed study time and random or staggered patient entry, a clinical trial in which patients are lost to follow-up, or a clinical trial in which there are multiple causes of failure but interest centers on only one of them. In all of these studies, interest focuses on estimating the underlying survival distribution of the time to occurrence of the primary event of interest.

In the above settings one of two primary methods of analysis is usually employed. First, a distribution-free estimator, suggested by Kaplan and Meier (1958) is often routinely used. Alternatively, maximum likelihood methods may be used to estimate the survival function when the survival distribution belongs to some parametric family.

Both of these approaches have some shortcomings. While the Kaplan-Meier nonparametric estimator has many desirable large sample properties, its small sample properties are suspect. In particular, it is biased for finite samples and the magnitude of this bias is inversely related to sample size (cf. Gross and Clark (1975)). Recently, Miller (1983) showed that the asymptotic efficiencies of the Kaplan-Meier product-limit estimator, relative to the maximum likelihood estimator of a parametric survival function, are low when the censoring proportions are high or for surviving fractions that are close to zero. For low surviving fractions the variance of the Kaplan-Meier estimator is large and any drop in efficiency represents a real loss of accurancy. Thus, Miller concludes that "parametric modelling should be considered as a means of increasing the precision in the estimation of small tail probabilities." In addition, the Kaplan-Meier estimator's performance is suspect for small samples as noted by Geurts (1985), for nonproportional hazard rate censoring, and by Chen, Hollander, and Langberg (1982) who state that, for proportional hazard censoring, "for small samples the Kaplan-Meier estimator is biased and non-negligibly so under heavy censoring at median to low values of the survival function."

If the underlying parametric survival distribution is known, then parametric maximum likelihood methods are preferred. However, in reality, one rarely has this knowledge available. This approach assumes a plausible parametric model and estimation is carried out by the method of maximum likelihood (c.f. Nelson (1982) or Bain (1978)). Typical models used are the exponential (McCool (1974)), the Weibull (Bain and Antle (1970)), the normal and log normal (Harter and Moore (1965)), the gamma (Engelhardt and Bain (1978)), the log-logistic (Bain, Eastman and Engelhardt (1973)), the Pareto (Proschan (1963)), and the exponential power distribution (Smith and Bain (1975)). In light of the fact that it is often difficult to distinguish between the candidate distributions (cf. Bain and Englehardt (1980), Siswadi and Quesenberry (1982), Kent and Quesenberry (1982)) the investigator may very well choose the incorrect distribution. If the wrong model is chosen, then we know the parametric maximum likelihood method will lead to asymptotically biased results and will perhaps be quite inappropriate. This problem of misspecification bias is perhaps the largest concern in using this method.

Recently, Moeschberger and Klein (1985) have described a method for improving the product-limit estimator's performance when the data are censored at some fixed time point. The approach involved estimating the survival function beyond the last death by an appropriately chosen parametric function. In this paper we present an extension of this method to randomly censored data. The proposed estimator treats the uncensored observations nonparametrically and uses a parametric model only for the censored observations. This new estimator retains most of the distribution-free properties of Kaplan-Meier and, yet, allows one to estimate the function with reasonable accuracy in the tails. In the next section the motivation and construction of the proposed estimator is discussed. In Section 3 properties of the estimator, when the correct parametric model is chosen, are presented. Results of a Monte Carlo robustness study which compares the proposed estimator with the Kaplan-Meier approach, a smoothed Kaplan-Meier approach, and the maximum likelihood approach are presented in Section 4. Finally, an example of estimating the survival distribution of transplant patients enrolled in The Ohio State University Bone Marrow Transplant Program is discussed in Section 5.

2. CONSTRUCTION OF THE ESTIMATOR

For the jth individual under observation or on test, let X_j denote the potential time to occurrence of the event of interest and let Y_j denote the potential time to censoring, $j = 1, \ldots, n$. In this discussion we shall make the common assumption that X_j, Y_j are independent with survival functions $\bar{F}(\cdot)$ and $\bar{G}(\cdot)$, respectively. We observe T_j = minimum (X_j, Y_j) and $\delta_j = 1$ if $X_j \le Y_j$ (a death) and $\delta_j = 0$ if $X_j > Y_j$ (a censored observation). Our goal is to estimate $\bar{F}(x) = P\,[X > x]$.

Suppose, that based on a preliminary graphical analysis of the data (cf. Elandt-Johnson and Johnson (1980), Nelson (1982)) or based on a theoretical model of the disease process, a plausible model for $\bar{F}(x)$ is selected to be $S(x|\underline{\theta})$, where $\underline{\theta}$ is a vector of unknown parameters. Let $\hat{\underline{\theta}}$ be a consistent estimator of $\underline{\theta}$ based on the censored sample (T_j, δ_j), $j = 1, \ldots n$.

The proposed estimator is constructed by analogy to the complete data problem where a natural estimator of $\bar{F}(x)$ is the proportion of (X_j, Y_j) pairs with $X_j > x$. In the censored data problem we observe points along the line $X = Y = T$ along with a ray of possible values of the unobservable coordinate (see Figure 1). For those pairs (T_j, δ_j) with $T_j > x$ (rays a and b in Figure 1) we are sure that the corresponding X_j is greater than x. For pairs (T_j, δ_j) with $T_j \le x$ and $\delta_j = 1$ (ray c) in Figure 1 we are positive that X_j is less than or equal to x. When $T_j \le x$ and $\delta_j = 0$ (ray d in Figure 1) we can not determine with certainty if the true unobservable X_j is greater than x or not. In this case an estimate of the chance of the true X_j being greater than x, in light of the observable information (T_j, δ_j), is $P(X_j > x | T_j = \tau_j, \delta_j = 0) = P(X_j > x | X_j > T_j, Y_j = T_j)$ which, using the assumed parametric model $S(\cdot|\underline{\theta})$ for $\bar{F}$, is $S(x|\hat{\underline{\theta}})/S(T_j|\hat{\underline{\theta}})$. This suggests the following estimator:

$$\hat{\bar{F}}(x) = \sum_{j=1}^{n} \phi_j\,(x|S(\cdot|\hat{\underline{\theta}}))/n \qquad (1)$$

$$\text{where} \quad \phi_j(x|S(\cdot|\hat{\underline{\theta}})) = \begin{cases} 1 \text{ if } T_j > x,\ \delta_j = 0 \text{ or } 1 \\ 0 \text{ if } T_j \le x,\ \delta_j = 1 \\ S(x|\hat{\underline{\theta}})/S(T_j|\hat{\underline{\theta}}) \text{ if } T_j \le x,\ \delta_j = 0 \end{cases} \qquad (2)$$

3. PROPERTIES OF THE ESTIMATOR WHEN THE MODEL IS CORRECT

In this section we assume, up to the unknown parameter $\underline{\theta}$, that $S(x) = \bar{F}(x)$ for all x. Before deriving properties of $\hat{\bar{F}}(x)$, in this case, we present an alternate expression for $\hat{\bar{F}}(x)$. Let $\chi(A)$ denote the indicator function of the event A and define

$$\hat{Q}_n(x) = \sum_{j=1}^{n} \chi(T_j > x)/n \text{ and } \hat{Q}_{on}(x) = \sum_{j=1}^{n} \chi(T_j \le x, \delta_j = 0)/n.$$

Then one can show that

$$\hat{\bar{F}}(x) = \hat{Q}_n(x) + \int_0^x \frac{S(x|\hat{\underline{\theta}})}{S(u|\hat{\underline{\theta}})}\, d\,\hat{Q}_{on}(u). \qquad (3)$$

This representation allows us to prove the following theorem (proof in the Appendix).

Theorem 1. Under the random censoring model with X_i, Y_i independent and $S(x|\underline{\theta}) = \bar{F}(x)$, $\hat{\bar{F}}(x) \to \bar{F}(x)$ uniformly in x with probability one if $\hat{\underline{\theta}} \to \underline{\theta}$ with probability one.

The representation (A.4) in the Appendix allows us to prove, by arguments very closely related to those in Breslow and Crowley (1974), the following weak convergence result.

Theorem 2. Assume the independent random censoring model with $S(x|\underline{\theta}) = \bar{F}(x)$. If $\sqrt{n}\ \hat{\underline{\theta}}$ is a consistent estimator of θ which converges in distribution to a normal random variable then $\sqrt{n}\ (\hat{\bar{F}}(x) - \bar{F}(x))$ converges weakly to a Gaussian process with mean 0.

While the proof (which we outline in the Appendix) of the above result is straight-forward, evaluation of the limiting covariance is difficult, especially for estimators of $\underline{\theta}$ obtained by iterative techniques, since this covariance involves the limiting covariance of $(n^{1/2}(\hat{Q}_n(x)-Q(x)),\ n^{1/2}(\hat{\theta}-\theta))$ and $(n^{1/2}(\hat{Q}_{on}(t)-Q_o(t)),\ n^{1/2}(\hat{\theta}-\theta))$. However in some special cases this limiting covariance can be obtained. One such special case is given below.

Corollary 1. If X_i and Y_i are independent exponential random variables with hazard rates λ, β respectively, and $S(x|\hat{\underline{\theta}}) = \exp(-x/\theta)$ in (1) and $\hat{\theta} = \Sigma T_i/d$ where d is the observed number of deaths ($\hat{\theta}$ is the maximum likelihood estimator of $\theta=1/\lambda$), then

$Z_n(x) = \sqrt{n}(\hat{\hat{F}}(x) - \exp(-x/\theta)) \rightarrow$ Gaussian process with mean 0

$$\text{and } V(Z(x)) = \begin{cases} (2\lambda^2x^2+\lambda x-2(1-\exp(-\lambda x))^2)\exp(-2\lambda x) & \text{if } \lambda = \beta \\ \left[\dfrac{\lambda(\exp-(\lambda+\beta)x)-\exp(-2\lambda x))}{(\lambda - \beta)}\right. & \\ \left. + \dfrac{\lambda(\lambda+\beta)(\beta^2x^2-(1-\exp(-\beta x))^2)\exp(-2\lambda x)}{\beta^2}\right] & \text{if } \lambda \neq \beta \end{cases} \quad (4)$$

The details of the proof of this corollary may also be found in the Appendix.

From Kaplan and Meier (1958) we note that in the exponential case the asymptotic variance of the product limit estimator is

$$\lambda\exp(-2\lambda x)(\exp((\lambda+\beta)x)-1)/(\lambda+\beta) \quad (5)$$

which is always greater than (4). Figure 2 shows a plot of the asymptotic relative efficiency ((5)/(4)) as a function of the censoring fraction $p = \beta/(\lambda + \beta)$ for $\lambda=1$, $0 \leq p \leq .5$ at the 10th, 50th and 90th percentile of the survival distribution. Note from this figure that the relative efficiency of our proposed estimator improves with increased censoring and increasing time. While this result is true for the exponential case our Monte Carlo study presented in the next section seems to indicate it is true for the Weibull case as well.

4. MONTE CARLO ROBUSTNESS STUDY

To study the performances of our estimator when the incorrect parametric form is chosen for S, a Monte Carlo study was performed. The study consisted of simulating 1,000 samples of size 25 or 50 with 10%, 30%

or 50% censoring from the following distributions: (In each case we fixed the mean life at 1 to make comparisons easier.)

1. E: Exponential;
2. $W(\alpha)$: Weibull with $S(t) = \exp(-\beta t^{\alpha})$;
3. $G(\alpha)$: Gamma with probability density function

$$f(t) = \alpha^{\beta} t^{\alpha-1} \exp(-\beta t)/\Gamma(\beta);$$

4. LN(s): Log normal with second moment equal to the second moment of a Weibull (α, β);
5. $N(\sigma)$: Normal with mean 1 and standard deviation σ;
6. $EP(\beta)$: Exponential power distribution (Smith & Bain (1975)) with

$$S(t) = \exp(1-\exp((t/\alpha)^{\beta}))$$

7. GOM (γ): Gomertz with $S(t) = \exp\ [\beta(1-\exp(\gamma t))/\gamma]$, β, γ, $t > o$;
8. $P(\alpha)$: Pareto with $S(t) = [\beta/(\beta+t)]^{\alpha}$;
9. BT(p): A bathtub shaped hazard distribution (Glaser (1980)) which is a mixture of an exponential with probability (1-p) and a gamma $(3,\lambda)$ with probability p; and
10. LL (β): Log logistic with $S(t) = (1+\gamma t^{\beta})^{-1}$, β, $\gamma > \beta$ o.

Censoring distributions were the exponential for all distributions. In addition, proportional hazards censoring was used for the $W(\alpha)$ and $EP(\beta)$. The distributions selected contain a variety of shapes for the hazard rate; decreasing hazard rates for the $W(\alpha)$, $\alpha < 1$, $G(\alpha)$, $\alpha<1$, $P(\alpha)$; constant hazard rates for E; increasing hazard rates for the $W(\alpha)$, $\alpha > 1$, $G(\alpha)$, $\alpha > 1$, $EP(\beta)$, $\beta > 1$; and U shaped hazard rates for the BT(p), $EP(\beta)$, $\beta < 1$.

Table 1 summarizes the results of the study by reporting the observed ratio of the mean squared error of the product-limit estimator to our suggested estimator based on either the exponential or Weibull distribution choice for S. For both models, parameters were estimated by the method of maximum likelihood. In this table a value greater than 1 implies that the product-limit estimator had a larger mean squared error.

Several conclusions can be drawn from the study. First, the estimator, based on S following a Weibull distribution, always has MSE less than or equal to that of the product-limit estimator. A similar result held true for the bias although these values are not shown. These differences seem to become more pronounced as t increases. The performance of the estimator based on S following an exponential distribution is not as clear-cut. It works better for those distributions with a decreasing or U shaped hazard rate, but performs poorer for increasing hazard rate distributions especially at large t. Our recommendation in light of this study is to use the Weibull choice for S when there is no obvious reason to choose another distribution.

Secondly, in all cases where our estimator had a smaller mean squared error it also had a smaller variance. This suggests that a conservative estimator of the variance of $\hat{\hat{F}}(t)$ is the estimated variance of the product limit estimator.

In an attempt to assess how much of the success of the proposed estimator may be due to smoothing, we compared the various estimators with a simple smoothing of a Kaplan-Meier by connecting the midpoints of the steps in the Kaplan-Meier estimator. Figures 3 through [illegible] plot relative MSE's of the smooth Kaplan-Meier, exponential maximum likelihood, Weibull maximum likelihood, exponential scores, and Weibull scores estimators of survival with 50 percent censoring versus the MSE of the unsmoothed Kaplan-Meier estimator for various typical distributions, for sample size n=50, and for 50 percent censoring. As one can see, the improvement of the Weibull scores estimator appears to be something in excess of just smoothing

As a measure of the overall performance of an estimator, we considered an estimator of the integrated mean squared error defined as

$$\text{IMSE}\ (\hat{\bar{F}}) = E\ \{\int_0^\infty (\hat{\bar{F}}(x) - \bar{F}(x))^2\ dx\},$$

where $\bar{F}(x)$ is the true survival function. We estimate the quantity by

$$\text{E}\hat{\text{I}}\text{MSE}\ (\hat{\bar{F}}) = \sum_{j=1}^{1000} \int_0^{x_{.95}} (\hat{\bar{F}}_j(x) - \bar{F}(x))^2\ dx/1000$$

where $F(x_p) = p$ and $\hat{\bar{F}}_j(x)$ is the estimator of survival on the j^{th} simulation. The ratio of this quantity for the Kaplan-Meier estimator to that of one of the other proposed specific estimators (smoothed Kaplan-Meier, exponential MLE, Weibull MLE, exponential scores, and Weibull scores) is reported in parentheses in the legend of each of the Figures 3-7. A value greater than one implies that the estimator under consideration performs better, on the whole, than the Kaplan-Meier estimator. These ratios can also be used to make other comparisons between the various estimators. For example, we note that the Weibull MLE and Weibull scores methods are consistently better than the Kaplan-Meier and smoothed Kaplan-Meier, although there is no clear winner between the Weibull MLE and Weibull scores method when the underlying survival distribution is not Weibull. Also, it can be seen that the exponential scores method is consistently better than the exponential MLE, however, neither of these methods is consistently better than Kaplan-Meier or its smoothed counterpart.

5. EXAMPLE

We illustrate the estimator using the times to death, post-transplant of 42 patients enrolled in The Ohio State University Bone Marrow Transplant Program. The data represents three years of patient accrual at the end of this Phase II trial. Censoring arises due to random entry into the study. The data is in Table 2. There were 13 deaths so that 69% of the patients were still alive at study's end. The estimate of the exponential parameter was $\hat{\lambda} = 9.38\text{x}10^{-4}$. The estimate of the Weibull parameters were $\hat{\alpha} = .0033$, $\hat{\beta} = .7895$ which suggests a rapidly decreasing hazard rate. This is to be expected since graft versus host disease and other complications tend to kill patients early, if at all.

Figure 8 is a plot of the Kaplan-Meier estimator and the new estimator based on both a Weibull and an exponential choice of S. Notice that all three estimators agree until the censored observations begin to appear. For the newly proposed estimators the jump sizes at each death is 1/n while the Kaplan-Meier estimator has random jump sizes increasing with time.

Acknowledgments

We wish to express our appreciation to Dr. Peter Tutschka of The Ohio State University for providing the data in Table 2. This research was supported by the U.S. Air Force Office of Scientific Research under contract AFOSR-82-0307.

References

Bain, L.J. (1978). Statistical Analysis of Reliability and Life Test Models. Marcel-Dekker, New York.

Bain, L.J., and Antle, C. E. (1970). Inferential procedures for the Weibull and generalized gamma distributions. Aerospace Research Laboratories, Air Force Systems Command, USAF, Wright-Patterson AFB, Ohio. ARL-70-0266.

Bain, L.J., Eastman, J., and Englehardt, M. (1973). A study of life-testing models and statistical analysis for the logistic distribution. Aerospace Research Laboratories, Air Force Systems Command, USAF, Wright-Patterson AFB, Ohio. ARL-73-0009.

Bain, L.J. and Englehardt, M. (1980). Probability of correct selection of Weibull versus gamma based on likelihood ratio. Communications in Statistics - Theory and Methods. A 9(4): 375-381.

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimator under random censorship. Annals of Statistics 2:437-453.

Chen, Y.Y., Hollander, M., and Langberg, N.A. (1982). Small sample results for the Kaplan-Meier estimator. Journal of the American Statistical Association 67:141-144.

Elandt-Johnson, R. and Johnson, N. (1980). Survival Models and Data Analysis, Wiley, New York.

Englehardt, M., and Bain, L.J. (1978). Construction of optimal unbiased inference procedures for the parameters of the gamma distribution. Technometrics 20:485-489.

Geurts, J.H.J. (1985). Some small sample non-proportional hazard results for the Kaplan-Meier estimator. Statistica Neerlandica 39:1-13.

Glaser, R. (1980). Bathtub and related failure rate characterizations. Journal of the American Statistical Association 75:667-672.

Gross, A. and Clark, V. (1975). Survival Distributions: Reliability Applications in the Biomedical Sciences, Wiley, New York.

Harter, H. L. and Moore, A. H. (1965) Maximum likelihood estimation of the parameters of gamma and Weibull populations from complete and from censored samples. Technometrics 7:639-643.

Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 53:457-481.

Kent, J. and Quesenberry, C.P. (1982). Selecting among probability distributions used in reliability. Technometrics 24:59-65.

McCool, J.I. (1974). Inferential techniques for Weibull populations. Wright-Patterson Air Force Base, Ohio, ARL Technical Report ARL-74-0180.

Mendenhall, W. and Lehman, E.H., Jr. (1960). An approximation to the negative moments of the positive binomial useful in life testing. Technometrics 2:227-242.

Miller, R.G. (1983). What price Kaplan-Meier? Biometrics 39:1077-1083.

Moeschberger, M. and Klein, J. (1985). A comparison of several methods of estimating the survival function when there is extreme right censoring. Biometrics 41:253-260.

Nelson, W. (1982). Applied Life Data Analysis. Wiley, New York.

Proschan, F. (1963). Theoretical explanation of observed decreasing failure rate. Technometrics 5:375-384.

Siswadi, and Quesenberry, C.P. (1982). Selecting among Weibull, lognormal, and gamma distributions using complete and censored samples. Naval Research Logistics Quarterly 29:557-569.

Smith, R. and Bain, L. (1975). An exponential power life-testing distribution. Communications in Statistics 4(5):469-481.

Appendix - Proof of Theorem 1

First note that $\hat{Q}_n(x) \to Q(x) = \bar{F}(x)H(x)$ uniformly in x with probability one (w.p.1)

$\hat{Q}_{on}(x) \to Q(x) = \int_o^x \bar{F}(u)(-dH(u))$ uniformly

in x w.p.1; and

$$\frac{S(x|\hat{\underline{\theta}})}{S(u|\hat{\underline{\theta}})} \to \frac{\bar{F}(x)}{\bar{F}(u)} \quad \text{almost surely (a.s.).}$$

Also note that $\bar{F}(x) = Q(x) + \int_o^x \frac{\bar{F}(x)}{\bar{F}(u)} dQ_o(u)$. (A.1)

Let $Y_{1n}(x) = [\hat{Q}_n(x) - Q(x)]$, (A.2)

$Y_{2n}(x) = [\hat{Q}_{on}(x) - Q_o(x)]$,

and $W_n(x,u) = \dfrac{S(x|\hat{\underline{\theta}})}{S(u|\hat{\underline{\theta}})} - \dfrac{\bar{F}(x)}{\bar{F}(u)}$. (A.3)

It follows that

$$\hat{\bar{F}}(x) - \bar{F}(x) = Y_{1n}(x) + \int_o^x W_n(x,u)dQ_o(u)$$

$$+ \int_o^x \frac{\bar{F}(x)}{\bar{F}(u)} dY_{2n}(u)$$

$$+ \int_o^x W_n(x,u)dY_{2n}(u). \quad \text{(A.4)}$$

Since $\hat{\bar{F}}(x) - \bar{F}(x)$ is a continuous function of $Y_{1n}(x)$, $Y_{2n}(x)$, and $W_n(x,u)$ in the sup norm and each of these processes converge to 0 uniformly, the uniform consistency of $\hat{\bar{F}}(x)$ follows.

Outline of the Proof of Theorem 2:

From A.4 it follows that

$$\sqrt{n}(\hat{\bar{F}}(x)-\bar{F}(x))=\sqrt{n}Y_{1n}(x)+\int_0^x \sqrt{n}W_n(x,u)dQ_0(u)+\sqrt{n}Y_{2n}(x)$$

$$-\bar{F}(x)\int_0^x \sqrt{n}Y_{2n}(u)d(\bar{F}^{-1}(u))+\int_0^x \sqrt{n}W_n(x,u)dY_{2n}(u) \qquad (A.5)$$

$=A_n(x)+B_n(x)+C_n(x)-D_n(x)+R_n(x)$, say. Expanding $W_n(x,u)$ as a function of $\hat{\theta}$ in a Taylor Series about θ yields

$$\sqrt{n}\, W_n(x,u) = \sqrt{n}\,(\hat{\theta}-\theta)d[S_0(x|\theta)/S_0(u|\theta)]/d\theta+o_p(1). \qquad (A.6)$$

So $\sqrt{n}\, W_n(x,u)$ converges to $W(x,u)$, a Gaussian Process in x and u. Also $\sqrt{n}Y_{1n}(x)$, $\sqrt{n}Y_{2n}(x)$ converge weakly to a bivariate Gaussian process, $(Y_1(x), Y_2(x))$. To prove limiting normality of $\sqrt{n}(\hat{\bar{F}}(x)-\bar{F}(x))$ note that A_n, B_n, C_n and D_n all converge weakly in the supremum metric to Gaussian processes A,B,C and D, respectively and R_n converges a.s. to 0 in this metric.

Evaluation of the limiting Variance of Z(x) in the Exponential Case
(needed for Corollary 1)

First note that $\sqrt{n}\,(\hat{\theta}-\theta)$ converges in distribution to a normal random variable with mean 0 and variance $(\lambda+\beta)/\lambda^3$. By A.6 we have

$$\sqrt{n}\, W_n(x,u)= \sqrt{n}\,(\hat{\theta}-\theta)\,(x-u)\exp\{-(x-u)/\theta\}/\theta^2+o(1), \qquad (A.7)$$

hence

$$\mathrm{Cov}(W(x,u),W(y,r))=\lambda(\lambda+\beta)\,(x-u)\,(y-r)\,\exp\{-\lambda(x+y-u-r)\}. \qquad (A.8)$$

Now by A.7

$$Cov(\sqrt{n}\, Y_{1n}(x), \sqrt{n}\, W_n(y,u)) = n\,(y-u)\lambda^2 \exp\{-\lambda(y-u)\} Cov(\widehat{Q}_n(x), \widehat{\theta}). \qquad (A.9)$$

To evaluate $Cov(\widehat{Q}_n(x),\widehat{\theta})$ we have

$$Cov(\widehat{Q}_n(x),\widehat{\theta}) = E[\widehat{Q}_n(x)\widehat{\theta}] - E[\widehat{Q}_n(x)]\, E[\widehat{\theta}]$$

$$= E[\quad \chi(T_1>x)\widehat{\theta}] - Q(x)\, E[\Sigma T_i]\, E[d^{-1}]$$

$$= E[\widehat{\theta}|T_1>x]\, Q(x) - n\, Q(x)\, E[d^{-1}]/(\lambda+\beta)$$

$$= \left\{(E[T_1|T_1>x] + E[\sum_{2}^{n} T_i]) - n/(\lambda+\beta)\right\} \cdot Q(x) E[d^{-1}]$$

$$= (x + 1/(\lambda+\beta) + (n-1)/(\lambda+\beta) - n/(\lambda+\beta)) \cdot Q(x) E[d^{-1}]$$

$$= x\, Q(x) E[d^{-1}] \qquad (A.10)$$

From Mendenhall and Lehman(1960) we note that

$$E[d^{-1}] \approx (n-2)/\{n[(n-1)\lambda/(\lambda+\beta)]\}. \qquad (A.11)$$

Combining A.9, A.10 and A.11 and taking the limit as n tends to infinity we obtain

$$n\; Cov(Y_1(x), W(y,u)) = \lambda\,(\lambda+\beta)\,(y-u)\, x \exp\{-[\,(\lambda+\beta)\, x + \lambda(y-u)]\}\,. \qquad (A.12)$$

A similar argument is used to show that

$$n\, Cov(Y_2(x), W(y,u)) =$$
$$\lambda\beta(y-u)\exp\{-\lambda(y-u)\}\{-x\exp\{-(\lambda+\beta)x\} + (1-\exp\{-(\lambda+\beta)\, x\})/(\lambda+\beta)\}. \qquad (A.13)$$

Routine calculations yield the following (for $x < y$):

$$n\; Cov(Y_1(x), Y_1(y)) = \exp\{-(\lambda+\beta)y\} - \exp\{-(\lambda+\beta)\,(x+y)\}, \qquad (A.14)$$

$$n\,\mathrm{Cov}(Y_2(x),Y_2(y)) = \beta[\lambda+\beta\exp\{-(\lambda+\beta)y\}-\lambda\exp\{-(\lambda+\beta)x\}$$

$$-\beta\exp\{-(\lambda+\beta)(x+y)\}]/(\lambda+\beta)^2, \qquad (A.15)$$

and

$$n\mathrm{Cov}(Y_1(x),Y_2(y)) = -\beta[-\exp\{-(\lambda+\beta)x\}+\exp\{-(\lambda+\beta)(x+y)\}]/(\lambda+\beta). \qquad (A.16)$$

From the representation A.5 it follows that

$$V(Z(x)) = V(A(x))+V(B(x))+V(C(x))+V(D(x))+2\mathrm{Cov}(A(x),B(x))$$
$$+2\mathrm{Cov}(A(x),C(x)) -2\mathrm{Cov}(A(x),D(x))+2(\mathrm{Cov}(B(x),C(x)) \qquad (A.17)$$
$$-2\mathrm{Cov}(B(x),D(x))-2\mathrm{Cov}(C(x),D(x)).$$

From A.14 we have

$$V(A(x)) = \exp\{-(\lambda+\beta)x\}-\exp\{-2(\lambda+\beta)x\}; \qquad (A.18)$$

from A.8

$$V(B(x))=2\int_0^x\int_0^y \mathrm{Cov}(W(x,r),W(x,y))dQ_0(r)dQ_0(y)$$

$$=\lambda(\lambda+\beta)\exp\{-2\lambda x\}[\exp\{-\beta x\}+\beta x-1]^2/\beta^2; \qquad (A.19)$$

from A.15

$$V(C(x))=\beta[\lambda+(\beta-\lambda)\exp\{-(\lambda+\beta)x\}-\beta\exp\{-2(\lambda+\beta)x\}]/(\lambda+\beta)^2; \qquad (A.20)$$

from A.15

$$V(D(x))=2\int_0^x\int_0^y \mathrm{Cov}(Y_2(r),Y_2(y))\lambda^2\exp\{-\lambda(x-r+x-y)\}\,dr\,dy$$

$$=2\beta\lambda^2\exp\{-2\lambda x\}[-(\lambda+\beta)^2/(2\beta\lambda(\lambda-\beta))+\exp\{2\lambda x\}/(2\lambda)$$
$$+(\beta^2+\lambda^2)\exp\{x(\lambda-\beta)\}/(\lambda\beta(\lambda-\beta))-(\lambda+\beta)\exp\{\lambda x\}/(\beta\lambda) \qquad (A.21)$$
$$+(\lambda+\beta)\exp\{-\beta x\}/(\beta\lambda)-\exp\{-2\beta x\}/(2\beta)]/(\lambda+\beta)^2, \text{ if } \beta\neq\lambda$$
$$=\exp\{-2\lambda x\}[\exp\{2\lambda x\}/4+\lambda x-\exp\{\lambda x\}+\exp\{-\lambda x\}-\exp\{-2\lambda x\}/4], \text{ if } \beta=\lambda;$$

from A.12

$$\mathrm{Cov}(A(x),B(x))=\int_0^x \mathrm{Cov}(Y_1(x),W(x,u)\, dQ_0(u)$$

$$=(\lambda+\beta)\lambda x\exp\{-(2\lambda+\beta)x\}[(\beta x-1)+\exp\{-\beta x\}]/\beta; \qquad (A.22)$$

from A.16

$$\mathrm{Cov}(A(x),C(x))=-\beta[-\exp(-(\lambda+\beta)x\}+\exp\{-2(\lambda+\beta)x\}]/(\lambda+\beta); \qquad (A.23)$$

from A.16

$$\mathrm{Cov}(A(x),D(x))=\int_0^x \mathrm{Cov}(Y_1(x),Y_2(u))\lambda\exp\{-\lambda(x-u)\}du \qquad (A.24)$$

$$=-\beta\exp\{-(\lambda+\beta)x\}/(\lambda+\beta)+\exp\{-(2\lambda+\beta)x\}-\lambda\exp\{-2(\lambda+\beta)x\}/(\lambda+\beta);$$

from A.13

$$\mathrm{Cov}(B(x),C(x))=\int_0^x \mathrm{Cov}(Y_2(x),W(x,u))dQ_0(u) \qquad (A.25)$$

$$=\lambda\exp(-\lambda x)[\beta x-1+\exp\{-\beta x\}]\{-x\exp\{-(\lambda+\beta)x\}+(1-\exp\{-(\lambda+\beta)x\})/(\lambda+\beta)\};$$

from A.13

$$\mathrm{Cov}(B(x),D(x))=\int_0^x\int_0^y \mathrm{Cov}(W(x,y),Y_2(r))\beta\lambda\exp\{-\lambda(x-y)\}\exp\{-(\lambda+\beta)r\}drdy$$

$$+\int_0^x\int_0^y \mathrm{Cov}(W(x,r),Y_2(y))\beta\lambda\exp\{-\lambda(x-y)\}\exp\{-(\lambda+\beta)r\}drdy$$

$$=\exp\{-2(\lambda+\beta)x\}[\lambda^2(\beta x+1)/\beta^2+\lambda^2/(\beta(\lambda+\beta))]$$

$$+\exp\{-(\lambda+\beta)x\}(\lambda/(\lambda+\beta))+\exp\{-\lambda x\}(\beta x-1)(\lambda/(\lambda+\beta))$$

$$+\exp\{-(2\lambda+\beta)x\}[(\beta x-1)\lambda^2/(\beta(\lambda+\beta))+\lambda^2x^2-2\lambda^2/\beta^2-\lambda/\beta]$$

$$-\exp\{-2\lambda x\}(\beta x-1)[\lambda^2/\beta^2+\lambda^2/(\beta(\lambda+\beta))+(\lambda/(\lambda+\beta)]; \qquad (A.26)$$

and from A.15

$$\mathrm{Cov}(C(x),D(x))=\int_0^x \mathrm{Cov}(Y_2(x),Y_2(u))\lambda\exp\{-\lambda(x-u)\}du$$

$$=\beta\lambda/(\lambda+\beta)^2-\lambda\exp\{-\lambda x\}/(\lambda+\beta)+(\beta^2+\lambda^2)\exp\{-(\lambda+\beta)x\}/(\lambda+\beta)^2$$

$$-\beta\exp\{-(2\lambda+\beta)x\}/(\lambda+\beta)+\beta\lambda\exp\{-2(\lambda+\beta)\}/(\lambda+\beta)^2. \qquad (A.27)$$

Substituting A.18-A.27 into A.17 we obtain (4) after some very tedious simplifications.

Table 1. Ratio of MSE of product limit estimator to MSE of new estimator for various distributions, percent censored, and percentiles of survival.

			$F(x_{.1})=.1$				$F(x_{.5})=.5$				$F(x_{.9})=.9$			
			n=25		n=50		n=25		n=50		n=25		n=50	
Distr. of Deaths	Distr. of Losses	Percent censored	Exp	Wie	Exp	Wie	Exp	Wie	Exp	Wie	Exp	Wie	Exp	Wie
E	E	10	1.01	1.01	1.01	1.01	1.05	1.04	1.05	1.04	1.22	1.18	1.22	1.18
		30	1.04	1.02	1.03	1.02	1.18	1.17	1.15	1.15	2.08	1.79	1.95	1.65
		50	1.09	1.06	1.09	1.05	1.33	1.29	1.29	1.28	3.85	2.96	4.34	2.74
W(1/2)	E	10	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.02	1.26	1.25	1.20	1.24
		30	1.00	1.00	1.00	1.00	1.05	1.08	1.03	1.06	2.10	2.34	1.20	2.55
		50	1.01	1.01	1.01	1.01	1.00	1.18	.97	1.19	5.99	5.64	3.59	7.59
W(2)	E	10	1.03	1.03	1.04	1.03	1.10	1.06	1.12	1.07	1.03	1.12	.96	1.12
		30	1.11	1.09	1.08	1.10	1.40	1.22	1.36	1.20	.70	1.48	.48	1.46
		50	1.21	1.17	1.13	1.19	1.63	1.38	1.51	1.36	.47	2.07	.29	1.90
W(4)	E	10	1.05	1.06	1.01	1.05	1.15	1.07	1.14	1.06	.85	1.10	.70	1.09
		30	1.00	1.16	.84	1.19	1.53	1.21	1.39	1.22	.34	1.32	.20	1.31
		50	1.02	1.34	.74	1.30	1.81	1.33	1.29	1.37	.18	1.57	.10	1.54
W(8)	E	10	1.04	1.07	.95	1.06	1.17	1.06	1.15	1.06	.68	1.09	.56	1.08
		30	.86	1.22	.62	1.21	1.42	1.18	1.36	1.19	.26	1.27	.14	1.25
		50	.78	1.38	.51	1.39	1.59	1.28	1.14	1.31	.14	1.55	.07	1.42
W(1/2)	W(1/2)	10	1.01	1.01	1.01	1.01	1.03	1.04	1.03	1.04	1.19	1.19	1.13	1.16
		30	1.04	1.02	1.04	1.02	1.08	1.15	1.02	1.15	1.65	1.65	1.09	1.65
		50	1.10	1.06	1.10	1.06	1.08	1.29	1.08	1.30	2.87	2.91	1.56	2.86
W(2)	W(2)	10	.99	1.00	1.01	1.01	1.04	1.04	1.05	1.04	.67	1.03	.90	1.17
		30	1.04	1.03	1.04	1.03	1.21	1.15	1.18	1.17	.64	1.74	.46	1.81
		50	1.09	1.06	1.08	1.06	1.33	1.30	1.16	1.30	.63	2.98	.36	2.95
W(4)	W(4)	10	1.01	1.01	1.01	1.01	1.05	1.04	1.05	1.05	.68	1.20	.55	1.16
		30	1.04	1.03	1.04	1.03	1.16	1.16	1.07	1.14	.27	1.73	.17	1.75
		50	1.09	1.05	1.09	1.06	1.00	1.31	.73	1.28	.23	2.68	.13	2.89
W(8)	W(8)	10	1.01	1.00	1.01	1.01	1.05	1.04	1.03	1.04	.51	1.18	.38	1.17
		30	1.04	1.02	1.04	1.02	1.05	1.15	.89	1.15	.20	1.76	.09	1.60
		50	1.10	1.06	1.10	1.06	.84	1.30	.53	1.30	.17	2.91	.09	2.89
G(1/2)	E	10	1.00	1.00	1.00	1.00	1.02	1.02	1.02	1.03	1.24	1.20	1.23	1.21
		30	1.00	1.00	1.01	1.00	1.06	1.10	1.07	1.10	2.63	1.88	1.72	1.66
		50	1.01	1.01	1.01	1.01	1.09	1.24	1.00	1.22	6.78	3.37	4.00	3.45
G(2)	E	10	1.03	1.02	1.03	1.02	1.09	1.05	1.07	1.05	1.13	1.12	1.06	1.13
		30	1.10	1.07	1.12	1.09	1.30	1.19	1.28	1.19	1.23	1.54	.93	1.48
		50	1.23	1.17	1.17	1.16	1.57	1.34	1.45	1.36	1.20	2.50	.76	2.15
G(4)	E	10	1.05	1.05	1.05	1.05	1.11	1.06	1.11	1.06	1.01	1.12	.93	1.10
		30	1.12	1.15	1.03	1.14	1.37	1.19	1.35	1.18	.65	1.36	.46	1.29
		50	1.24	1.27	1.10	1.28	1.43	1.26	1.38	1.32	.39	1.80	.26	1.61
G(8)	E	10	1.06	1.06	1.06	1.07	1.15	1.07	1.14	1.07	.89	1.08	.78	1.06
		30	1.12	1.21	.90	1.21	1.43	1.18	1.38	1.21	.41	1.26	.25	1.20
		50	1.23	1.45	.91	1.45	1.51	1.29	1.22	1.27	.24	1.59	.14	1.39

Table 1 (continued)
Page 2

Distr. of Deaths	Distr. of Losses	Percent censored	$F(x_{.1})=.1$ n=25 Exp	n=25 Wie	n=50 Exp	n=50 Wie	$F(x_{.5})=.5$ n=25 Exp	n=25 Wie	n=50 Exp	n=50 Wie	$F(x_{.9})=.9$ n=25 Exp	n=25 Wie	n=50 Exp	n=50 Wie
LN(.37)	E	10	1.05	1.07	1.03	1.07	1.12	1.06	1.13	1.07	.93	1.08	.81	1.04
		30	1.15	1.23	.92	1.23	1.41	1.19	1.30	1.15	.45	1.18	.29	1.14
		50	1.16	1.47	.91	1.48	1.40	1.18	1.06	1.15	.27	1.39	.15	1.31
LN(.51)	E	10	1.05	1.05	1.06	1.06	1.10	1.05	1.09	1.05	1.03	1.10	.92	1.06
		30	1.16	1.18	1.08	1.18	1.36	1.17	1.27	1.14	.68	1.23	.49	1.23
		50	1.30	1.37	1.13	1.40	1.47	1.22	1.27	1.25	.46	1.68	.26	1.47
LN(.61)	E	10	1.05	1.04	1.04	1.04	1.09	1.05	1.09	1.05	1.10	1.09	1.05	1.07
		30	1.13	1.16	1.10	1.16	1.33	1.18	1.27	1.17	.97	1.34	.75	1.28
		50	1.29	1.31	1.26	1.33	1.49	1.23	1.26	1.22	.69	1.74	.41	1.51
EP(.25)	E	10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.23	1.25	1.04	1.18
		30	1.00	1.00	1.00	1.00	1.02	1.04	1.02	1.04	1.79	2.18	1.07	2.28
		50	1.00	1.00	1.00	1.00	.88	1.18	.77	1.16	8.11	4.71	6.56	5.80
EP(.5)	E	10	1.00	1.00	1.00	1.00	1.02	1.03	1.02	1.03	1.32	1.27	1.26	1.22
		30	1.00	1.00	1.01	1.00	1.06	1.10	1.08	1.12	2.53	1.77	2.00	1.60
		50	1.01	1.01	1.02	1.01	1.10	1.25	1.05	1.26	7.26	2.70	4.71	2.47
EP(1)	E	10	1.01	1.01	1.02	1.01	1.07	1.05	1.06	1.05	1.17	1.19	1.09	1.16
		30	1.06	1.03	1.06	1.03	1.27	1.20	1.26	1.22	1.25	1.65	1.02	1.53
		50	1.12	1.07	1.13	1.06	1.53	1.35	1.57	1.43	1.35	1.94	.96	1.77
EP(8)	E	10	1.03	1.07	.95	1.06	1.17	1.07	1.17	1.06	.58	1.13	.46	1.11
		30	.80	1.21	.52	1.18	1.55	1.16	1.32	1.19	.22	1.37	.11	1.39
		50	.79	1.39	.45	1.35	1.46	1.23	.95	1.30	.10	1.77	.06	1.70
EP(.25)	EP(.25)	10	1.01	1.01	1.01	1.01	1.03	1.04	1.01	1.05	1.17	1.22	1.10	1.19
		30	1.04	1.02	1.03	1.02	1.00	1.15	.92	1.15	1.47	1.66	.98	1.51
		50	1.09	1.04	1.10	1.04	.90	1.32	.84	1.26	2.77	2.07	1.41	1.93
EP(.5)	EP(.5)	10	1.01	1.01	1.01	1.01	1.04	1.05	1.03	1.03	1.23	1.20	1.23	1.19
		30	1.04	1.02	1.04	1.02	1.12	1.17	1.11	1.14	2.17	1.66	1.71	1.58
		50	1.09	1.04	1.10	1.05	1.16	1.32	1.13	1.34	4.89	2.05	3.19	1.91
EP(1)	EP(1)	10	1.01	1.01	1.01	1.01	1.06	1.05	1.05	1.04	1.20	1.21	1.12	1.20
		30	1.04	1.02	1.04	1.02	1.20	1.16	1.18	1.15	1.15	1.57	.98	1.56
		50	1.08	1.04	1.08	1.04	1.37	1.32	1.36	1.33	1.47	2.29	1.04	1.86
EP(8)	EP(8)	10	1.01	1.01	1.01	1.01	1.06	1.03	1.02	1.04	.53	1.22	.38	1.20
		30	1.04	1.02	1.04	1.02	1.06	1.16	.88	1.15	.17	1.54	.10	1.57
		50	1.11	1.06	1.09	1.04	.75	1.29	.53	1.28	.15	2.05	.06	1.80
LL(2)	E	10	1.02	1.02	1.02	1.02	1.05	1.04	1.05	1.04	1.11	1.05	1.11	1.06
		30	1.10	1.08	1.09	1.08	1.22	1.15	1.17	1.12	1.59	1.36	1.48	1.28
		50	1.23	1.19	1.20	1.17	1.46	1.28	1.34	1.29	2.23	2.26	2.01	1.87
LL(3)		10	1.04	1.04	1.04	1.04	1.08	1.05	1.10	1.07	1.03	1.03	1.02	1.03
		30	1.13	1.15	1.08	1.15	1.32	1.16	1.27	1.16	.82	1.19	.59	1.08
		50	1.26	1.31	1.09	1.28	1.39	1.21	1.33	1.30	.57	1.75	.39	1.40
LL(4)		10	1.07	1.06	1.02	1.05	1.10	1.06	1.11	1.06	.92	1.01	.85	1.00
		30	1.08	1.20	1.03	1.20	1.43	1.21	1.27	1.14	.57	1.15	.35	1.00
		50	1.17	1.38	.92	1.37	1.51	1.25	1.15	1.17	.34	1.48	.19	1.15

Table 1 (continued)
Page 3

Distr. of Deaths	Distr. of Losses	Percent censored	F(x.1)=.1 n=25 Exp	F(x.1)=.1 n=25 Wie	F(x.1)=.1 n=50 Exp	F(x.1)=.1 n=50 Wie	F(x.5)=.5 n=25 Exp	F(x.5)=.5 n=25 Wie	F(x.5)=.5 n=50 Exp	F(x.5)=.5 n=50 Wie	F(x.9)=.9 n=25 Exp	F(x.9)=.9 n=25 Wie	F(x.9)=.9 n=50 Exp	F(x.9)=.9 n=50 Wie
P(2)	E	10	1.01	1.00	1.01	1.00	1.02	1.02	1.03	1.03	1.24	1.17	1.19	1.13
		30	1.03	1.02	1.03	1.02	1.11	1.11	1.12	1.13	2.21	1.84	1.74	1.71
		50	1.07	1.04	1.07	1.05	1.23	1.23	1.24	1.26	5.79	4.14	3.86	3.93
P(3)	E	10	1.01	1.01	1.01	1.01	1.03	1.03	1.04	1.04	1.25	1.17	1.21	1.15
		30	1.03	1.02	1.03	1.02	1.13	1.13	1.13	1.14	2.26	1.78	2.02	1.73
		50	1.07	1.04	1.08	1.05	1.25	1.24	1.27	1.27	6.04	3.67	4.85	3.74
P(5)	E	10	1.01	1.01	1.01	1.01	1.04	1.03	1.04	1.04	1.26	1.19	1.22	1.16
		30	1.03	1.02	1.04	1.02	1.15	1.14	1.15	1.15	2.18	1.72	2.11	1.73
		50	1.07	1.04	1.08	1.05	1.28	1.25	1.30	1.30	5.78	3.41	5.60	3.40
BT(.15)	E	10	1.01	1.01	1.01	1.01	1.04	1.03	1.04	1.04	1.26	1.20	1.24	1.19
		30	1.94	1.03	1.04	1.02	1.14	1.14	1.14	1.14	2.06	1.75	2.10	1.76
		50	1.09	1.06	1.08	1.05	1.29	1.27	1.27	1.28	4.95	3.32	4.98	3.14
BT(.25)	E	10	1.01	1.00	1.01	1.01	1.04	1.04	1.05	1.04	1.24	1.19	1.24	1.19
		30	1.03	1.02	1.04	1.02	1.15	1.14	1.15	1.15	2.20	1.83	2.23	1.83
		50	1.08	1.05	1.08	1.05	1.31	1.29	1.30	1.30	4.33	3.00	5.33	3.42
BT(.40)	E	10	1.01	1.01	1.01	1.01	1.05	1.04	1.05	1.04	1.27	1.23	1.23	1.19
		30	1.04	1.02	1.04	1.02	1.16	1.15	1.16	1.15	1.94	1.75	1.97	1.74
		50	1.08	1.05	1.08	1.05	1.34	1.30	1.32	1.32	3.48	2.69	3.64	2.59
BT(.6)	E	10	1.01	1.01	1.01	1.01	1.06	1.05	1.06	1.05	1.22	1.19	1.18	1.17
		30	1.04	1.02	1.04	1.02	1.23	1.19	1.21	1.18	1.55	1.58	1.34	1.48
		50	1.09	1.05	1.10	1.05	1.48	1.36	1.40	1.36	1.91	1.99	1.66	1.86
GOM(.5)	E	10	1.02	1.01	1.02	1.01	1.07	1.05	1.07	1.05	1.17	1.18	1.16	1.19
		30	1.05	1.03	1.05	1.03	1.25	1.19	1.23	1.19	1.33	1.58	1.22	1.58
		50	1.11	1.07	1.11	1.06	1.45	1.34	1.45	1.37	1.69	2.30	1.11	1.86
GOM(1)	E	10	1.02	1.01	1.02	1.01	1.09	1.06	1.09	1.06	1.09	1.18	.99	1.17
		30	1.06	1.03	1.06	1.03	1.31	1.21	1.28	1.18	.91	1.52	.67	1.50
		50	1.14	1.07	1.12	1.07	1.66	1.43	1.63	1.46	.76	1.74	.47	1.54
GOM(2)	E	10	1.04	1.02	1.02	1.01	1.15	1.09	1.15	1.09	.92	1.21	.75	1.19
		30	1.07	1.04	1.07	1.03	1.43	1.22	1.46	1.26	.52	1.52	.29	1.34
		50	1.14	1.08	1.07	1.07	1.80	1.38	1.76	1.44	.33	1.63	.18	1.32
N(.05)		10	.97	1.12	.86	1.11	1.18	1.08	1.17	1.08	.58	1.05	.39	1.03
		30	.70	1.36	.45	1.34	1.41	1.21	1.20	1.22	.16	1.13	.09	1.09
		50	.66	1.71	.36	1.50	1.23	1.29	.86	1.35	.09	1.22	.04	1.11
N(.10)		10	1.00	1.10	.93	1.10	1.18	1.07	1.14	1.06	.64	1.06	.47	1.04
		30	.78	1.35	.56	1.34	1.49	1.18	1.15	1.16	.20	1.15	.11	1.13
		50	.72	1.65	.41	1.60	1.38	1.27	.87	1.27	.11	1.35	.06	1.16
N(.15)		10	1.02	1.09	.90	1.08	1.18	1.07	1.17	1.08	.67	1.06	.53	1.07
		30	.88	1.27	.60	1.32	1.50	1.18	1.30	1.21	.24	1.21	.12	1.15
		50	.86	1.55	.51	1.53	1.41	1.31	1.01	1.28	.12	1.39	.06	1.24

Table 2. Times to death (in days post-transplant) of patients in OSU Bone Marrow Transplant Program (+ denotes still alive).

2, 27, 32+, 43+, 50, 55+, 62, 82+, 102+, 103+, 122, 145+, 148, 158, 162, 194+, 250+, 251, 267+, 276, 284+, 292+, 319+, 326+, 346+, 365+, 404+, 417, 418, 423+, 438+, 491, 584+, 595+, 613+, 642+, 649+, 693+, 707+, 746+, 755+, 826+

Figure 1.

Four possible rays in the censored data situation.

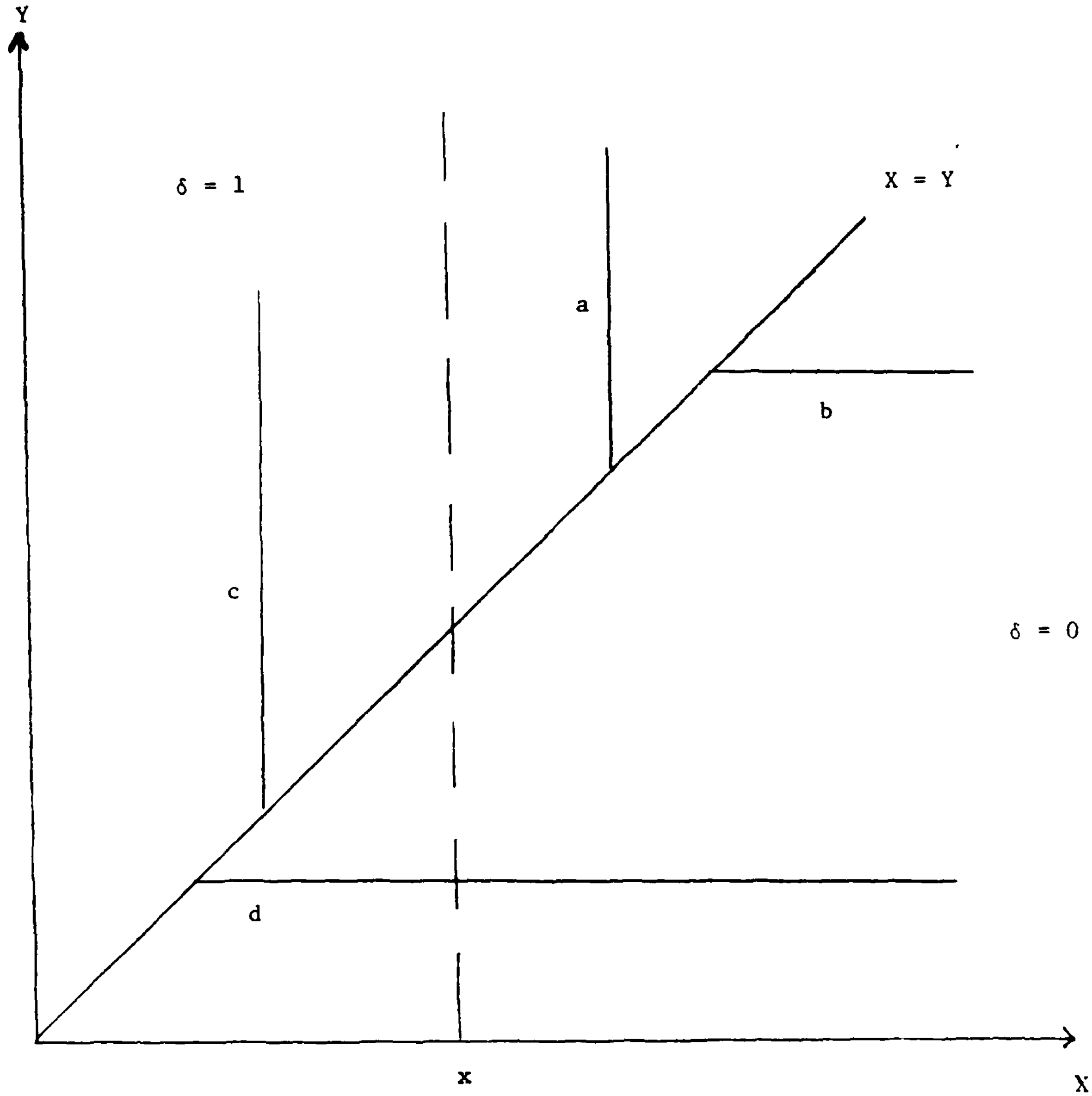


Figure 2.

Asymptotic Relative Efficiency (A.R.E.) as a function of the censoring fraction at the 10th (0), 50th (Δ), and 90th (+) percentile of $\bar{F}(x)$.

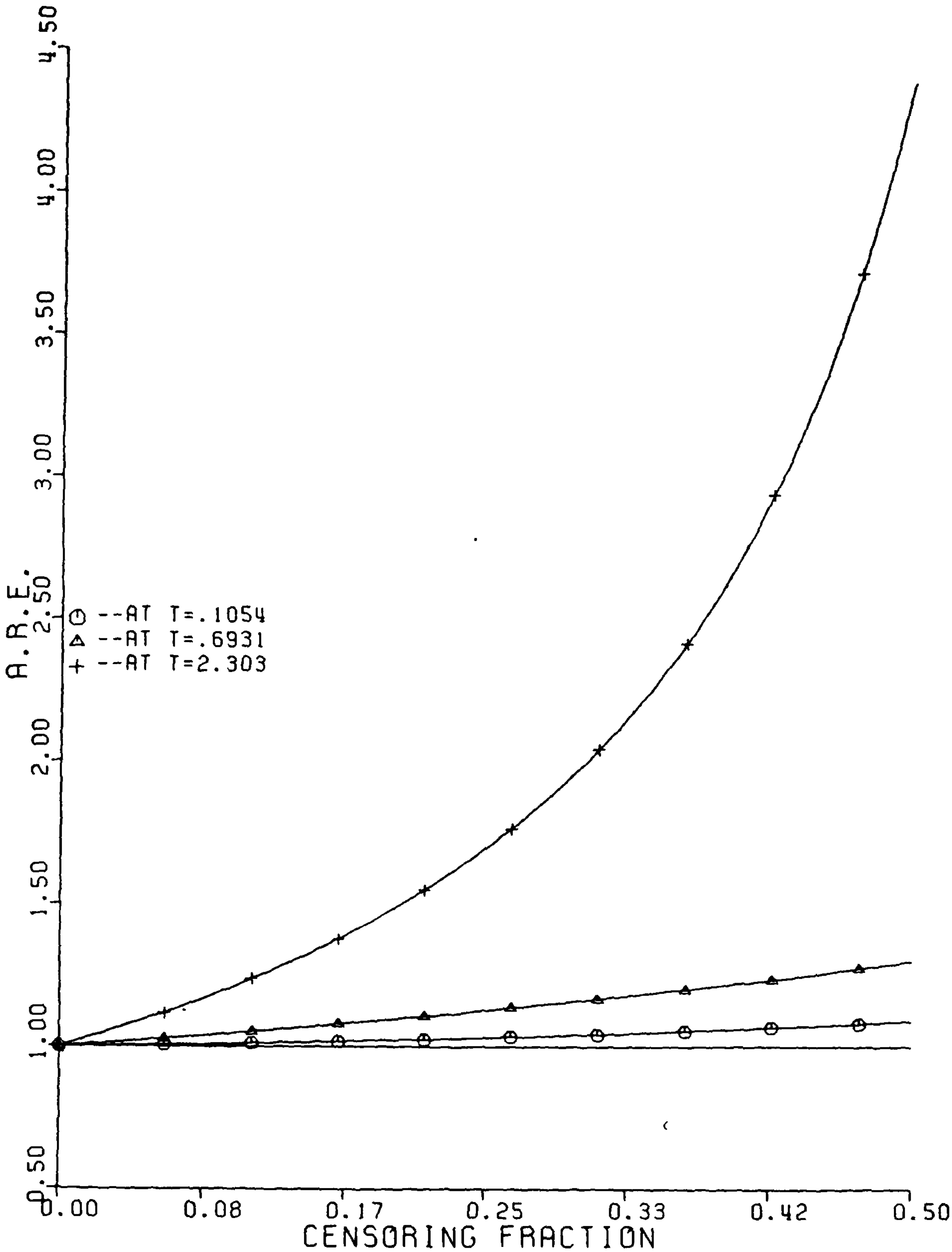


Figure 3.

Relative MSE of Estimators of Survival for the Weibull Distribution with 50% Censoring, n=50, α=0.5

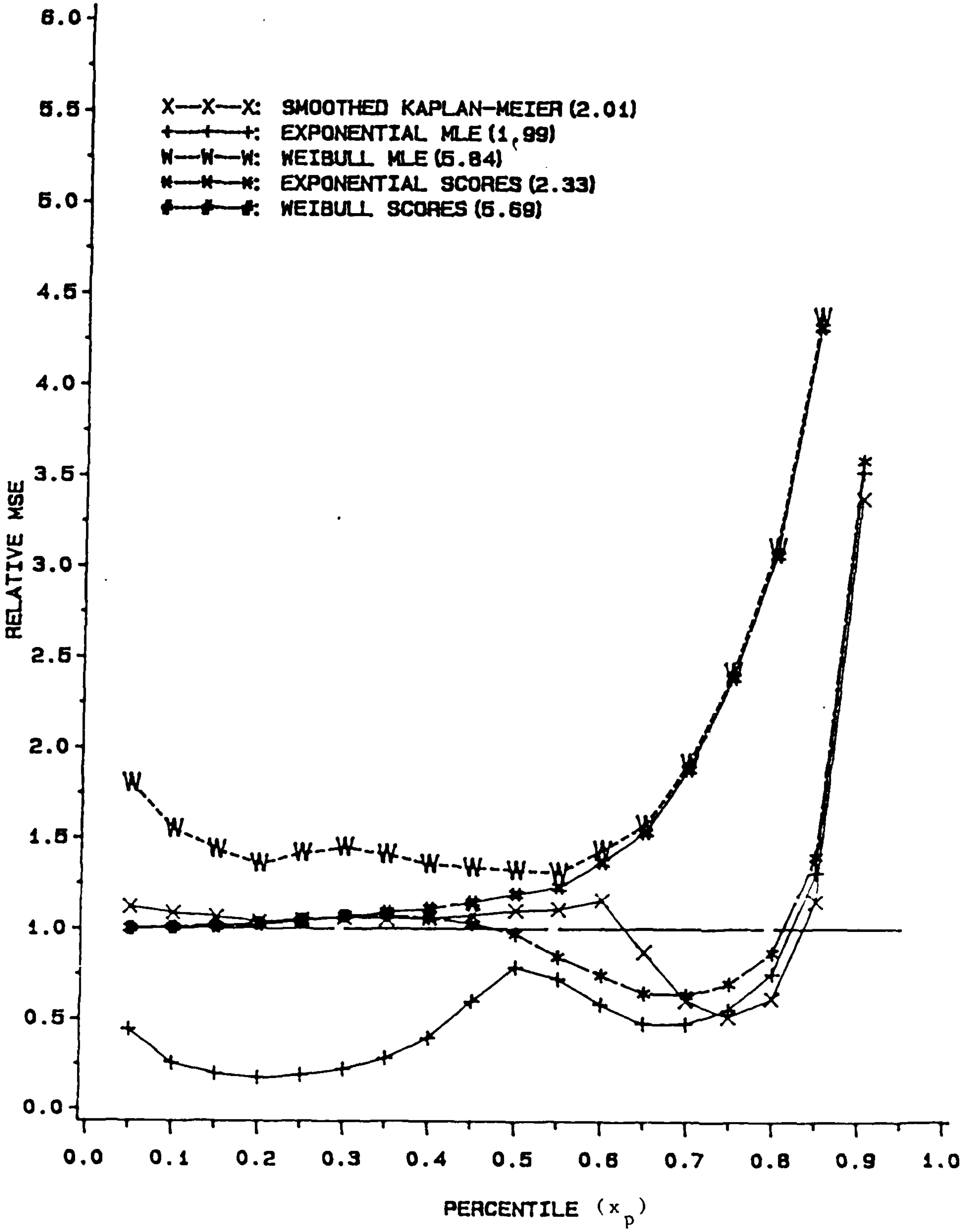


Figure 4.

Relative MSE of Estimators of Survival for the Exponential Power Distribution with 50% Censoring, n=50, β=0.5

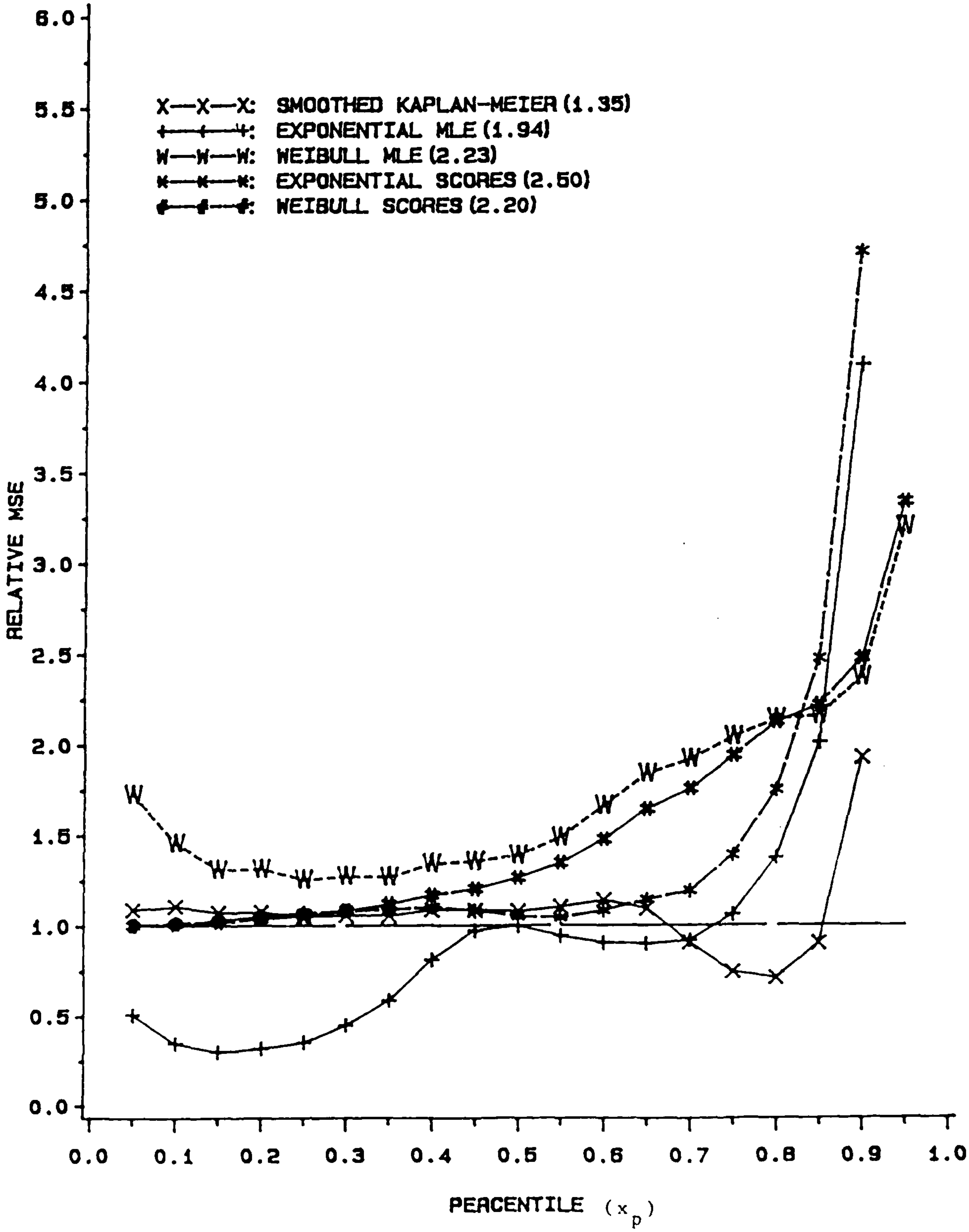


Figure 5.

Relative MSE of Estimators of Survival for the Log Normal Distribution with 50% Censoring, n=50, S=.6096

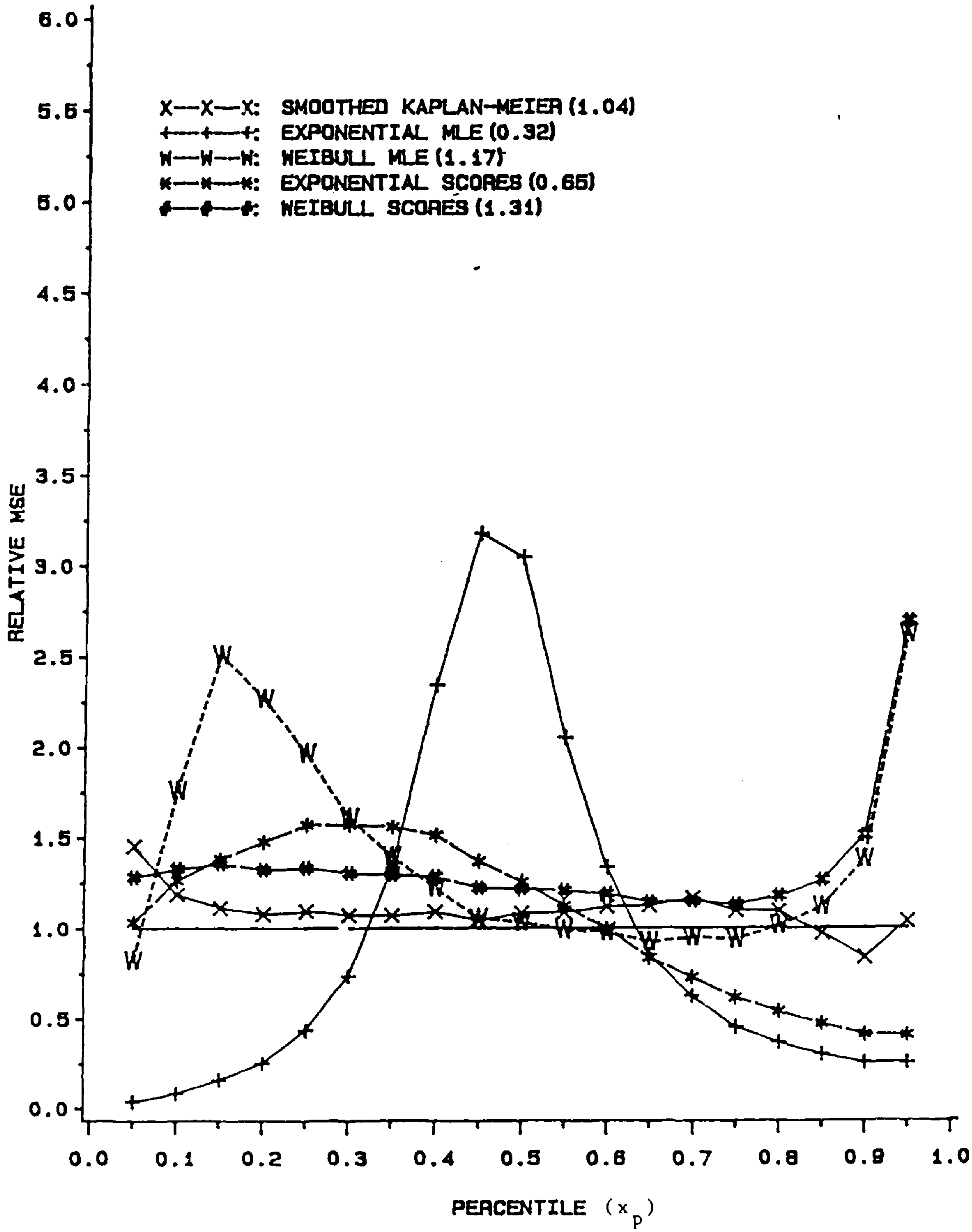


Figure 6.

Relative MSE of Estimators of Survival for the Gompertz Distribution with 50% Censoring, n=50, γ=1.0

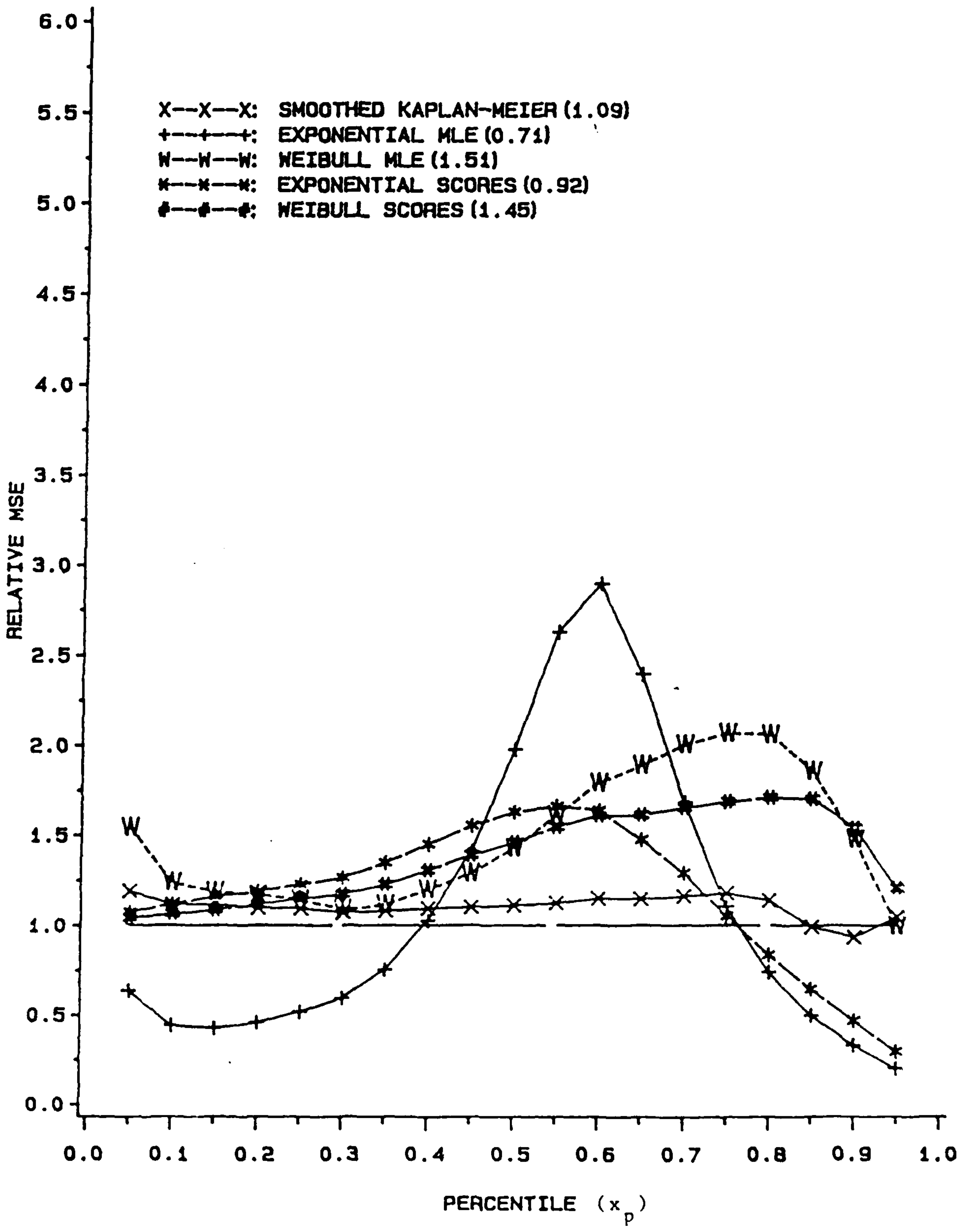


Figure 7.

Relative MSE of Estimators of Survival for the
Log Logistic Distribution with 50% Censoring, n=50, β=3.0

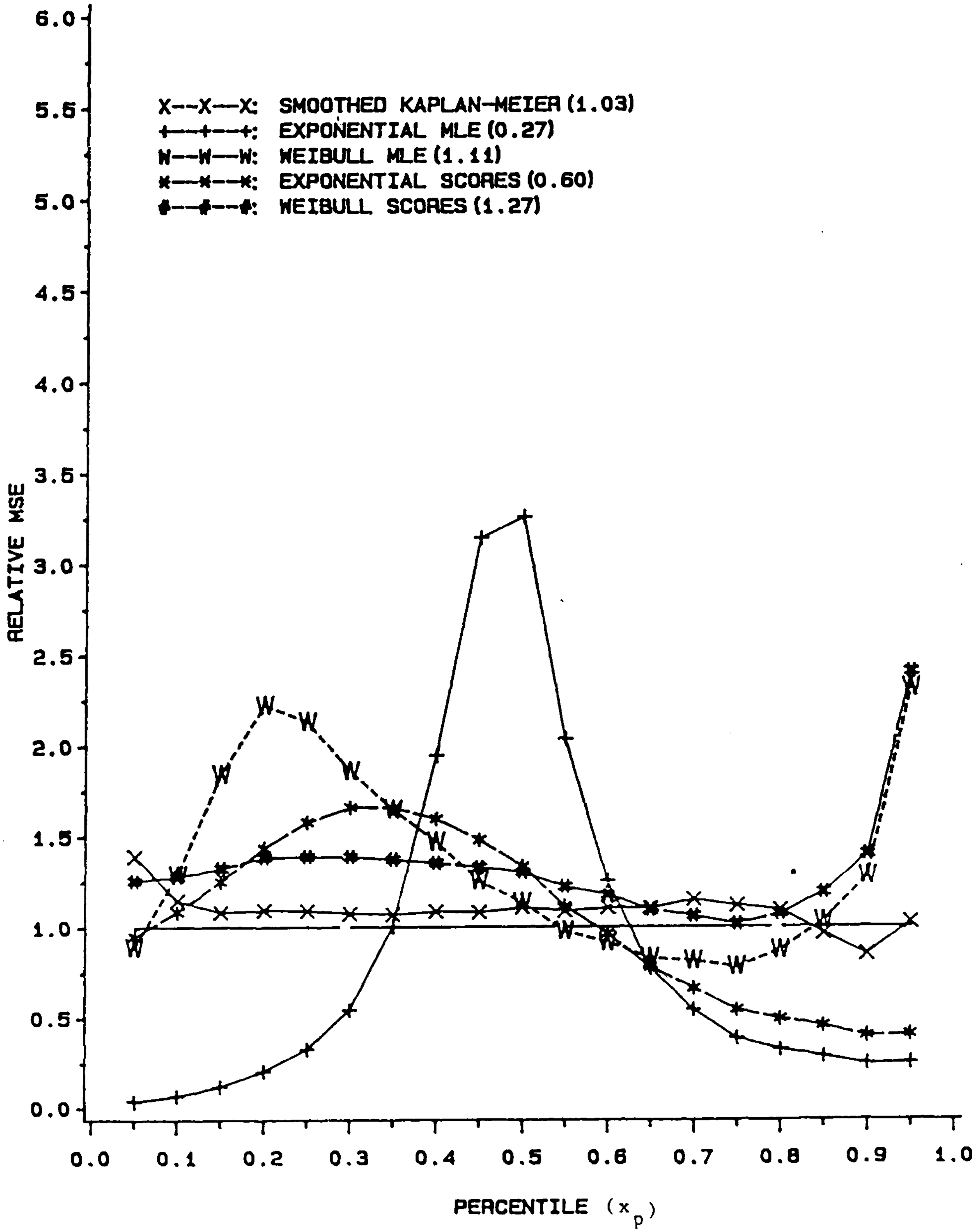
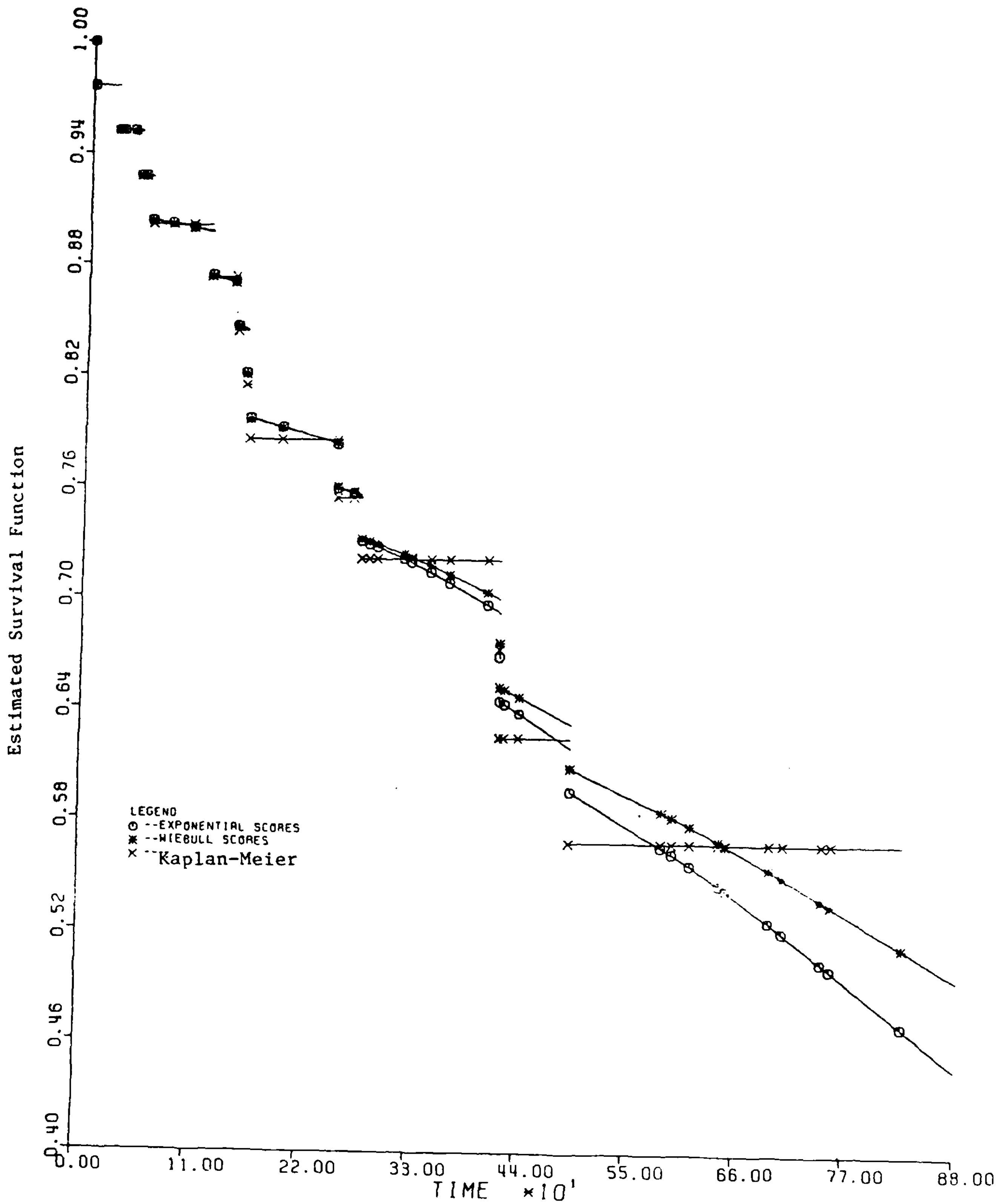


Figure 8.

Estimated survival function for Kaplan-Meier and proposed estimates for OSU transplant data



APPENDIX I

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY NA 2b. DECLASSIFICATION/DOWNGRADING SCHEDULE NA	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public Release; Distribution Unlimited
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION The Ohio State University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NM
6c. ADDRESS (City, State and ZIP Code) 1314 Kinnear Road Columbus, Ohio 43212	7b. ADDRESS (City, State and ZIP Code) Bldg. 410 Bolling AFB, DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-82-0307
8c. ADDRESS (City, State and ZIP Code) Bldg. 410 Bolling AFB, DC	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO.: 6.1102F PROJECT NO.: 2304 TASK NO.: WORK UNIT NO.:

11. TITLE (Include Security Classification)
Bivariate Models with a Random Environmental Factor

12. PERSONAL AUTHOR(S)
Sukhoon Lee and John P. Klein

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Final	FROM 9-1-82 TO 12-31-87	May 31, 1988	18

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	Series system, Reliability, Survival functions, Random environmental factor, Bivariate models
XXXXXXXXXXXXXXX			

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

In this paper a two component series system is modeled with a bivariate model that introduces a random environmental factor. Often an assumption is made that the two components of the system function independently. If this assumption is incorrectly made, predictions of system reliability may be very poor. Several authors have proposed models for dependence based on the assumption that the effect of exposure of a system to the environment is to simultaneously degrade (or improve) all components in the system. In this paper we propose a general model for a system affected by random environment and describe its properties. We restrict the model to one where the environmental effect is modeled by a gamma distribution. Finally, an example is presented.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION	
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
Brian W. Woodruff, Major	(202) 767-5027	AFOSR/NM

BIVARIATE MODELS WITH A RANDOM ENVIRONMENTAL FACTOR

Sukhoon Lee
Temple University
Philadelphia, PA USA

John P. Klein
The Ohio State University
Columbus, OH USA

1. INTRODUCTION

Consider a two component series system functioning in some environment. A common, though untestable, assumption is that the two components of the system function independently. If this assumption is incorrectly made predictions of the system reliability may be very poor (c.f. Klein and Moeschberger (1984)). Several authors have proposed models for dependence based on the assumption that the effect of exposure of a system to the environment is to simultaneously degrade (or improve) all components in the system (Hutchinson(1981), Bagchi and Samanta(1984), Lindley and Singpurwalla(1984), Oakes(1982), and Hougaard(1985)). In Section 2 we propose a general model for a system affected by a random environment and describe its properties. In Section 3 we restrict the model to one where the environmental effect is modeled by a gamma distribution and study its properties. Finally in Section 4, we conclude with an example.

2. GENERAL MODEL

We assume that under controlled conditions, as one may encounter in the testing or design stage of development, the times to failure of the two components, to be linked in a system, are X_0 and Y_0 with survival functions F_0, G_0, respectively. The two components are linked into a system and are put into operation under operating conditions. Suppose that under such conditions the effect of the environment is to degrade or improve each component by the same random amount, Z (with some distribution function H), which changes the marginal survival functions of the two components to F_0^Z and G_0^Z. A value of Z less than one means that component reliabilities are simultaneously improved, while a value of Z greater than one implies a joint degradation. We assume that the two components in a system under fixed conditions (i.e. given Z) function independently. The resulting joint survival function of the two components' lifetimes, (X,Y) in the operating environment is $F(x,y)=E\,(F_0^Z(x)\cdot G_0^Z(y))$. Denoting the cummulative hazard functions of X_0, Y_0 by $Q_{x0}(\cdot)$ and $Q_{y0}(\cdot)$, we have

$$F(x,y) = E[\exp\{-(Q_{x0}(x)+Q_{y0}(y))Z\}]\,. \tag{2.1}$$

Now we will discuss the properties of the model in terms of it's dependence structure, and reliability.

Property 1. The random variables X, Y of the lifetimes of the two components in a system under the operating environment are totally positive of order 2 dependent. (See Barlow and Proschan (1981) for a discussion of TP2 dependence.)

proof) Let $f_o(x)$ and $g_o(y)$ be the density function of X_o and Y_o respectively and f(x,y) be the joint density funtion of (X,Y). Define $q_{xo}(x) = f_o(x) / F_o(x)$. Then

$$f(x,y)=E\{Z^2F_o^{Z-1}(x)G_o^{Z-1}(y)f_o(x)g_o(y)\} \quad =q_{xo}(x)q_{yo}(y)\int z^2F_o^{z}(x)G_o^{z}(y)dH(z). \text{ So}$$

$$f(x_1,y_1)\cdot f(x_2,y_2) = \int\int_{u>v} q_{xo}(x_1)q_{yo}(y_1)\, q_{xo}(x_2)q_{yo}(y_2)\, u^2\, v^2F_o^{u}(x_1)G_o^{u}(y_1)\cdot F_o^{v}(x_2)G_o^{v}(y_2)dH(u)dH(v)$$

$$+ \int\int_{v>u} q_{xo}(x_1)q_{yo}(y_1)\, q_{xo}(x_2)q_{yo}(y_2)\, u^2\, v^2F_o^{u}(x_1)G_o^{u}(y_1)\cdot F_o^{v}(x_2)G_o^{v}(y_2)dH(u)dH(v)$$

$$=\int\int_{u>v} q_{xo}(x_1)q_{yo}(y_1)\, q_{xo}(x_2)q_{yo}(y_2)\, u^2\, v^2F_o^{u}(x_1)G_o^{u}(y_1)\, F_o^{v}(x_2)G_o^{v}(y_2)dH(u)dH(v)$$

$$+ \int\int_{u>v} q_{xo}(x_1)q_{yo}(y_1)\, q_{xo}(x_2)q_{yo}(y_2)\, u^2\, v^2F_o^{v}(x_1)G_o^{v}(y_1)\, F_o^{u}(x_2)G_o^{u}(y_2)dH(u)dH(v).$$

In the same manner, the other product $f(x_1,y_2)\cdot f(x_2,y_1)$ also can be written as integrals over the region u>v and $f(x_1,y_1)\cdot f(x_2,y_2) - f(x_1,y_2)\cdot f(x_2,y_1)$ can be written as

$$\int\int_{u>v} u^2\, v^2[\, q_{xo}(x_1)\, q_{xo}(x_2)\, F_o^{u}(x_1)\, F_o^{v}(x_2) \; - q_{xo}(x_1)\, q_{xo}(x_2)F_o^{v}(x_1)\, F_o^{u}(x_2)]\cdot [\; q_{yo}(y_1)\, q_{yo}(y_2)\, G_o^{u}(y_1)\, G_o^{v}(y_2) - q_{yo}(y_1)\, q_{yo}(y_2)G_o^{v}(y_1)\, G_o^{u}(y_2)]\, dH(u)dH(v)$$

The blanketed terms in the integrand are always nonnegative over the region u>v since $F_o^{v}(x) / F_o^{u}(x) = \exp[\, -(v-u)\, Q_{xo}(x)\,]$ is increasing in x for any F_o. Hence $f(x_1,y_1)\cdot f(x_2,y_2) - f(x_1,y_2)\cdot f(x_2,y_1) \geq 0$ for all $x_1<x_2$, $y_1 < y_2$, which leads to TP2 dependence. Q.E.D.

Since TP2 dependence implies that X is stochastically increasing in Y we obtain the following property.

Property 2. Under the same setting as in Property 1, the conditional hazard rates $q(x \mid Y=y)$ and $q(x \mid Y > y)$ are decreasing in y.

The property implies that the longer one component functions, the more reliable the other component in the system becomes.

From a different point of view we derive an inequality in terms of the conditional hazard rates which reflects the positive dependence of the model.

Property 3. Under the same setting as in Property 1, the model satisfies

$$q(x \mid Y = y) > q(x \mid Y > y).$$

Proof) Let $G_1(y)$ be the marginal survival function of y in the system exposed to the operating environment. Then $F(x \mid Y=y) = P(X > x \mid Y = y)$

$$= (\partial F(x, y)/\partial y) / (dG_1(y)/dy) = E(Z \cdot F_o^Z(x) \cdot G_o^Z(y)) / E(Z \cdot G_o^Z(y)).$$

Also $F(x \mid Y > y) = P(X > x \mid Y > y) = E(F_o^Z(x) \cdot G_o^Z(y)) / E(G_o^Z(y))$.

Hence, we obtain the following inequality,

$$\frac{q(x \mid Y = y)}{q(x \mid Y > y)} = \frac{E(Z^2 \cdot F_o^Z(x) \cdot G_o^Z(y))\, E(F_o^Z(x) \cdot G_o^Z(y))}{E^2(Z \cdot F_o^Z(x) \cdot G_o^Z(y))} \geq 1, \tag{2.2}$$

since $q(x \mid Y=y) = \partial[-\log F(x \mid Y=y)] / \partial x$

$$= E(Z^2 F_0^{Z-1}(x) \cdot G_0^Z(y) \cdot f_0(x)) / E(Z F_o^Z(x) \cdot G_o^Z(y))$$

$$= q_{xo}(x) E(Z^2 F_0^Z(x) \cdot G_0^Z(y)) / E(Z F_o^Z(x) \cdot G_o^Z(y)), \text{ and}$$

$$q(x \mid y > y) = q_{xo}(x) E(Z \cdot F_o^Z(x) \cdot G_o^Z(y)) / E(F_o^Z(x) \cdot G_o^Z(y)).$$

The inequality in (2.2) is obtained by Cauchy -Schwarz inequality and equality holds if and only if the random variable Z is constant. Q.E.D.

We note that this inequality should be contrasted with the notion of the quasi independence, which is defined by $q(x \mid Y = y) = q(x \mid Y > y)$. This is the necessary and sufficient condition that the marginal distributions under the dependent model can be recovered from the minimum of X, Y and the knowledge of which component caused the system to fail. In that case there exists a set of independent random variables which yields the same minimum and indicater of system failure as the dependent system. Futhermore these equivalent independent random variables have the same marginals as the dependent system (See Basu and Klein (1982)).

Up to now several properties have been explored in terms of the dependence structure induced by a random environment. Next we investigate the effects of the random environment on system reliability by comparing the reliability function with and without the random environmental effect. Conventional reliabilty theory commonly uses the knowledge of the component lifetimes and an assumption of independent component lifetimes in order to compute the system life distribution. In other words an investigator modeling system life, based on component information, may predict the reliability of the system, in our setting, with knowledge of $F_0(x)$ and $G_0(y)$ only by $R_{0S}(t) = F_0(t)\cdot G_0(t)$. The following theorem indicates how the two reliabilities are different in a series system.

Property 4. Suppose a two component system is serial, i.e., the system failsas soon as any one of the two components fails. Let $R_S(t)$ and $R_{0S}(t)$ denote the system reliabilities for the cases of a random environment and of a fixed environment.

i) If $E(Z) \leq 1$ then $R_S(t) \geq R_{0S}(t)$ for all t

ii) If $E(Z) > 1$ and $P(Z < 1) = 0$ then $R_S(t) < R_{0S}(t)$ for all t.

iii) If $E(Z) > 1$ and $P(Z > 1) > 0$ then there exists a t^* such that

$R_S(t) < R_{0S}(t)$ for all $t < t^*$ and $R_S(t) > R_{0S}(t)$ for all $t > t^*$.

Statement (iii) implies that even if the average operating environment is more severe than the controlled one, but there is a chance of better environment perhaps due to highly cautious maintenance, careful users, or effective usage, the system under a random environment becomes more reliable beyond a certain time.

Proof) The ratio of reliabilities for variable to fixed environment $R_{0S}(t)/R_S(t)$ is

$E\{\exp(-Q_0(t)Z)\} / \exp\{-Q_0(t)\}$ where $Q_0(t) = Q_{x0}(t)+Q_{y0}(t)$.

In the case of $E(Z) \leq 1$, $E\{\exp(-Q_0(t)Z)\} > \exp\{-Q_0(t)\cdot E(Z)\}$ by Jensen's inequality so i) follows immediately. Note that the equality holds if and only if Z=1 with probability 1.The statement (ii) follows by noting that $E\{\exp(-Q_0(t)Z)\} = \int \exp(-Q_0(t)z)dH(z) < \exp(-Q_0(t)z) \int dH(z) = \exp\{-Q_0(t)\}$.

To prove ~~(iii)~~, let $r(t) = E\{\exp(-Q_0(t)Z)\} / \exp\{-Q_0(t)\}$.

Then $r'(t)=q_0(t)E\{\exp(-Q_0(t)Z)\}\exp\{-Q_0(t)\}(1-s(t))$

where $s(t) = E\{Z\exp(-Q_0(t)Z)\}/ E\{\exp(-Q_0(t)Z)\}$ and $q_0(t)=dQ_0(t)/dt$.

Noting that $s(0)=E(Z)$ since $Q_0(0) = 0$, $E(Z)>1$ implies that $r'(0) < 0$. Since $r(t)$ is decreasing at $t = 0$ and $r(0) = 1$ this implies that $r(t) < 1$ for t in a neighborhood of $t=0$. To complete the proof it suffices to show that $r(t)$ is increasing beyond a certain point, which is true if $r'(t)$ is positive beyond that point. We claim $s(t)$ is decreasing in t and $s(t) < 1$ for large t under the given condition.Let us express $s(t)$ as

$$s(t)= \frac{E\{Z \exp(-Q_0(t)Z)\}}{E\{\exp(-Q_0(t)Z)\}} =\int zp(z|T>t)dz$$

where $p(z|T>t)\, dz = \dfrac{\exp(-Q_0(t)z)dH(z)}{c(Q_0(t))}$ and $c(Q_0(t)) = \int \exp(-Q_0(t)z)dH(z)$.

Noting that $p(z|T>t)$ is a density function, $s(t)$ can be expressed in terms of the conditional expecion $E(Z \mid T > t)$. Looking at the density $p(z|T>t)$ we see that

$$p(z|T>t_2)/ p(z|T>t_1) = c(Q_0(t_2))/ c(Q_0(t_1))\cdot \exp\{(Q_0(t_1) - Q_0(t_2))z\} \text{ for } t_1 < t_2$$

is decreasing in z. Then it is an immediate consequence of the following lemma, due to Lehmann that $E(Z| T > t)$ is decreasing in t

Lemma (Lehmann(1959),pg74) Let $p_\theta(x)$ be a family of densities on the real line with monotone likelihood ratio in x. If $\psi(x)$ is nondecreasing function of x, then $E_\theta(\psi(x))$ is a nondecreasing function of θ.

Let $\theta = 1/t$. Denote $p_\theta(z) = P(z|T>t)$. Then $p_\theta(z)$ has monotone likelihood ratio in z. So $E(Z)$ is nondecreasing in θ, which implies that $E(Z|T>t)$ is decreasing in t. Now it remains to be shown that $s(t) < 1$ for some $t > 0$. Let $p(z)=(z-1)\exp(-Q_0(t)\cdot z)$ and note that $p(0) = -1$ and $p(z)$ has maximum $[Q_0(t)\{\exp(Q_0(t)+1)\}]^{-1}$ at $z_0=(1+Q_0(t))\cdot Q_0^{-1}(t)$ and $p(z)$ is increasing for $z < z_0$ and decreasing $z > z_0$. Suppose $P(Z < 1) = \varepsilon > 0$.For any $0 < \delta < \varepsilon$, there exists a closed interval [u,v] contained in (0, 1) such that $A(u,v) = H(v) - H(u) \geq \delta$. Then $E(Z\exp(-Q_0(t)Z)) - E(\exp(-Q_0(t)Z))$

$$= \int_0^1 p(z)dH(z) + \int_0^\infty p(z)dH(z) \leq \int_u^v p(z)dH(z) + [Q_0(t)\cdot\exp(1+Q_0(t))]^{-1}\cdot A(1,\infty)$$

$$\leq p(v)\cdot A(u,v) + [Q_0(t)\cdot\exp(1+Q_0(t))]^{-1}\cdot A(1,\infty) < (v-1)\exp(-Q_0(t)v)\,\delta + [Q_0(t)\cdot\exp(1+Q_0(t))]^{-1}.$$

Since the last term is negative if and only if $(1-v)\,\delta > [e\cdot Q_0(t)\cdot\exp\{(1-v)Q_0(t)\}]$ *there exists a* t^* such that $E(Z\exp(-Q_0(t)Z)) - E(\exp(-Q_0(t)Z)) < 0$, that is, $s(t^*) < 1$. Q.E.D.

This property implies that methods which are based only on components' information (assuming independence) over estimate the reliability at an earlier stage ignoring potential failures from a harsh environment which may be encountered in the beginning stage under the operating condition, while under estimating the possible gains in reliability at later stage from a better environment which meets requirement of each system's susceptibility.

The proof of the property yielded an interesting result about the conditional distribution H( ) of a random environmental factor.

Property 5. The mean and variance of a random environmental factor Z among system which survive to a given time t, $E(Z \mid T > t)$ and $V(Z \mid T > t)$ are decreasing in t.

This property indicates that average environmental factor of the surviving systems declines with time since the systems under harsher environments tend to fail first. Also it is noted that the variability of environmental factor of the surviving systems is reduced with time.

We conclude this section by mentioning an curious phenomenon of the hazard rate. In the series system problem the life system distribution after incorporating a random environmental factor has hazard rate $q_s(t) = q_0(t)\cdot E(Z\exp(-Q_0(t)Z)) / E(\exp(-Q_0(t)Z))$. However $E(Z\exp(-Q_0(t)Z)) / E(\exp(-Q_0(t)Z))$ has been shown to be decreasing in t. Thus the lifetime distribution can often have a decreasing hazard rate which the variable environment may cause while the component hazard rates are not decreasing. One plausible explanation is that the population is subject to an early heavy selection of systems under most severe environments. This should be contrasted to reliability of system operating in a fixed environment where the systems may have a variety of shapes for the hazard rates.

3. THE MODEL WITH A GAMMA ENVIORNMENTAL FACTOR

In this section, we assume that the random environmental factor Z follows a gamma distribution with density function is $h(z) = \{\Gamma(\alpha)\}^{-1}\beta^{\alpha}\exp(-z/\beta)z^{\alpha-1}$, $\alpha > 0$ $\beta > 0$. This distribution is chosen because it is analytically tractable, and because it is flexible in that it allows a variety of shapes including the exponential and bell-shaped density. The joint survival function for (X, Y) is $F(x, y)= \beta^{\alpha} / \{\beta+Q_{xo}(x)+Q_{yo}(y)\}^{\alpha}$. (3.1)

and the marginal survival functions are $F_1(x) = \beta^{\alpha} / \{\beta+Q_{xo}(x)\}^{\alpha}$,

$G_1(y) = \beta^{\alpha} /\{\beta+Q_{yo}(y)\}^{\alpha}$.

Oakes (1982) proposed a similar model based on an extension of Clayton's (1978) model for bivariate lifetables. His model is derived by assuming that given w, X and Y are independent random variables with survival functions $\{\exp(-F_{ok}{}^{1-\theta}(x)+1)\}^{w}$ and $\{\exp(-G_{ok}{}^{1-\theta}(y)+1)\}^{w}$, respectively. Then assuming w has a gamma distribution with $\alpha = 1/(\theta-1)$ and $\beta = 1$ the joint survival function is $F(x, y) =[F_{ok}{}^{1-\theta}(x) + G_{ok}{}^{1-\theta}(y) - 1]^{1/(1-\theta)}$ where F_{ok}, and G_{ok} are the marginal survivals and $\theta > 1$.

For the Oakes' model the marginals under a fixed environment (i. e. given w) depend on the environmental parameter θ, while in our model they are free of the dependence measure α. The reverse situation holds under the random environment.

While the two joint survival functions appear quite different the basic marginal free structure is the same. That is both models have the same nonparametric dependence structure, the copula introduced by Sklar (1959) and studied by Schweizer and Wolfe (1981). Here the copula is $C(u,v) = [\ u^{-1/\alpha} + v^{-1/\alpha} - 1\ ]^{-\alpha}$, $\alpha=1/(\theta-1)$.

We list some properties of the gamma model we have obtained through the copula.

1) The probability of concordance is $(\alpha+1)/(2\alpha+1)$.

2) Since the copula C(u, v) depends only on α, only the shape parameter α affects dependence structure which is induced by the environment.

3) Since the copula C(u,v) is decreasing in α, and two variables are independent if and only if their copula is $u \cdot v$, the larger the shape parameter α is, the less the dependence is induced.

4) As α goes to 0 the copula converges to min(u,v) which is the copula of maximal positive association.

5) If we consider two environments characterized by gamma distributions $\alpha < \alpha'$ then X, Y under α is more positively associated than X,Y under α'.

We conclude this section with the following property of our model.

Property 6. The random environmental factor for those systems for which component A has functioned more than x time units and component B has functioned y time units also follows a gamma distribution with same shape parameter α and scale parameter $Q_{xo}(x) + Q_{yo}(y) + \beta$. While for the population of the systems whose components failed at time X=x, Y=y the environmental factor follows a gamma distribution with shape parameter $\alpha+2$ and scale parameter $Q_{xo}(x)+Q_{yo}(y)+\beta$.

Property 6 indicates that the mean of the environmental factor for the population of systems whose components are functioning at time t is a decreasing function of t. Another point to be noted from this property is that the density of the environmental factor for the population of the systems whose components' lifetimes have X > x, Y > y has the shape paramater α, which is identical to that in the unconditional density of Z. It can be interpreted that the dependence structure between the components of all the functioning system beyond a certain time t has the same as the dependence structure between components of a system operating at time 0.

4. EXAMPLES

As examples we consider the case when both components have a Weibull form parameters (η_1, λ_1) and (η_2, λ_2), respectively. That is, $F_0(x) = \exp(-\lambda_1 x^{\eta_1})$. The Weibull distribution, which may have increasing ($\eta > 1$), decreasing ($\eta < 1$) or constant failure rate ($\eta = 1$) has been shown experimentally to provide a reasonable fit to many different types of survival data. (See Bain (1978)). The resulting joint reliability of the two components' lifetimes, (X,Y) in the operating environment is $F(x,y) = E[\exp(-Z(\lambda_1 x^{\eta_1} + \lambda_2 y^{\eta_2}))]$. (4.1)

The model described above for a general distribution of the environmental stress has a particular dependence structure which we summarize in the following property.

Property 7. Let (X,Y) follow the model (4.1) where Z is a positive random variable with finite $\left(\frac{r}{\eta_1} + \frac{s}{\eta_2}\right)^{th}$ inverse moment. Then

$$E(X^r Y^s) = \lambda_1^{-r/\eta_1} \lambda^{-s/\eta_2} \Gamma(1 + r/\eta_1)\, \Gamma(1 + s/\eta_2)\, E(Z^{-(r/\eta_1 + s/\eta_2)}) \quad (4.1)$$

When the appropriate moments exist, we have

(A) $E(X) = E(X_0)\, E(Z^{-1/\eta_1})$,

(B) $V(X) = E(X_0^2)\, Var(Z^{-1/\eta_1}) + E(Z^{-1/\eta_1})^2\, Var(X_0)$,

(C) $Cov(X,Y) = E(X_0)\, E(Y_0)\, Cov(Z^{-1/\eta_1}, Z^{-1/\eta_2})$ which is strictly positive.

If $\eta_1 = \eta_2 = \eta$ then the correlation between (X,Y) is

$$\rho = \frac{\Gamma(1+1/\eta)^2\, Var(Z^{-1/\eta})}{Var(Z^{-1/\eta})\, \Gamma(1+2/\eta) + (\Gamma(1+2/\eta) - \Gamma(1+1/\eta)^2)\, E(Z^{-1/\eta})^2} \quad (4.2)$$

In this case the correlation is bounded above by $\Gamma(1+1/\eta)^2 / \Gamma(1+2/\eta)$. Figure 1 shows the maximal correlation as a function of η for $\eta \in (0, 10)$. Note that this maximal correlation is an increasing function of η.

Exact expressions for the quantities of interest can be computed when a particular model is assumed for the distribution of Z. We shall consider the gamma and uniform models. For the gamma model for Z, the joint survival function is

$$F(x,y) = \beta^\alpha \,/\, [\beta + \lambda_1 x^{\eta_1} + \lambda_2 y^{\eta_2}]^\alpha \quad (4.3)$$

which is a bivariate Burr Distribution (see Takahasi (1965)), the marginal distributions are univariate Burr distributions with

$$E(X) = (\lambda_1/\beta)^{-1/\eta_1} \Gamma(1+1/\eta_1)\Gamma(\alpha - 1/\eta_1)\, \Gamma(\alpha), \text{ if } \alpha > 1/\eta,$$

$$\mathrm{Var}(X) = (\lambda_1/\beta)^{-2/\eta_1} \left\{ \frac{\Gamma(1+2/\eta_1)\Gamma(\alpha-2/\eta_1)}{\Gamma(\alpha)} - \left[\frac{\Gamma(1+1/\eta_1)\Gamma(\alpha-1/\eta_1)}{\Gamma(\alpha)}\right]^2 \right\}, \text{ if } \alpha > 2/\eta_1$$

with similar expressions for E(Y), Var(Y). The covariance of (X,Y) is Cov(X,Y)

$$= (\lambda_1/\beta)^{-1/\eta_1} (\lambda_2/\beta)^{-1/\eta_2} \left\{ \frac{\Gamma(\alpha-1/\eta_1-1/\eta_2)}{\Gamma(\alpha)} - \frac{\Gamma(\alpha-1/\eta_2)\,\Gamma(\alpha-1/\eta_2)}{\Gamma(\alpha)} \right\} \cdot \Gamma(1+1/\eta_1)\,\Gamma(1+1/\eta_2)$$

for $\alpha > 1/\eta_1 + 1/\eta_2$. For the gamma model, the reliability function for a bivariate series system is given by

$$R_s(t) = (1+(\lambda_1/\beta)t^{\eta_1} + (\lambda_2/\beta)t^{\eta_2})^{-\alpha}, \qquad (4.4)$$

and for a parallel system by

$$R_p(t) = (1+(\lambda_1/\beta)t^{\eta_1} + (1+ (\lambda_2/\beta)t^{\eta_2})^{-\alpha} - (1+(\lambda_1/\beta)t^{\eta_2} + (\lambda_2/\beta)t^{\eta_2})^{-\alpha} \qquad (4.5)$$

Figure 2 is a plot of the series system reliability for $\lambda_1 = 1$, $\lambda_2 = 2$ and $\eta_1 = \eta_2 = 1$. The figure shows the reliablity for $\alpha = 1/2, 1, 2, 4$, and the independent Weibull model. In all cases, $\beta = 1$. For this figure we note that for fixed $\lambda_1, \lambda_2\ \eta_1, \eta_2$, t, the series system reliability is a decreasing function of the shape parameter α. Figure 3 is plot of the parallel system reliability (4.5) for the above parameters. Again, the reliability is a decreasing function of α. Also in both the series and parallel system reliability, the shape of the reliability function is quite different from that encountered under independence. Additional plots with other parameter values ,which can be found in Lee(1986) show similar survival curves.

The gamma model is often a reasonable model for the environmental stress as discussed in the previous section. However, in some cases, such as when the operating environment is always more severe than the laboratory environment, the support of H may be restricted to some fixed interval. A possible model for such an environmental stress is the uniform distribution over [a,b]. For this model, the joint survival function is

$$F(x,y) = \frac{[\exp(-b(\lambda_1 x^{\eta_1} + \lambda_2 y^{\eta_2})) - \exp(-a(\lambda_1 x^{\eta_1} + \lambda_2 y^{\eta_2}))]}{(b-a)(\lambda_1 x^{\eta_1} + \lambda_2 y^{\eta_2})} \tag{4.6}$$

$$E(X) = \lambda_1^{-1/\eta_1} \Gamma(1+1/\eta_1)\,\eta_1 (b^{(\eta_1-1)/\eta_1} - a^{(\eta_1-1)/\eta_1})/\{(\eta_1-1)(b-a)\} \quad \text{if } \eta_1 \neq 1,$$

$$= \log(b/a)/[\lambda_1(b-a)] \quad \text{if } \eta_1 = 1,$$

$$\mathrm{Var}(X) = \frac{\eta_1 \lambda_1^{-2/\eta_1}}{(b-a)} \{ \Gamma(1+2/\eta_1)\,\eta_1 (b^{(\eta_1-2)/\eta_1} - a^{(\eta_1-2)/\eta_1})$$

$$\frac{- \Gamma(1+1/\eta_1)^2 \eta_1 (b^{(\eta_1-1/\eta_1)} - a^{(\eta_1-1/\eta_1)})^2}{(\eta_1-1)^2(b-a)} \} \quad \text{if } \eta_1 \neq 1,2,$$

$$= 2/(\lambda_1^2 ab) - \log(b/a)^2/[(b-a)\lambda_1]^2 \quad \text{if } \eta_1 = 1,$$

$$= \lambda_1^{-1}[\frac{\log(b/a)}{(b-a)} - \frac{\pi}{(b^{1/2} + a^{1/2})^2}] \quad \text{if } \eta_1 = 2.$$

and

$$\mathrm{Cov}(X,Y) = \frac{\Gamma(1+1/\eta_1)\,\Gamma(1+1/\eta_2)}{\lambda_1^{1/\eta_1} \lambda_2^{1/\eta_1}} \{ \frac{\eta_1 \eta_2}{(\eta_1\eta_2 - \eta_1 - \eta_2)} \frac{(b^{\frac{\eta_1\eta_2 - \eta_1 - \eta_2}{\eta_1 \eta_2}} - a^{\frac{\eta_1\eta_2 - \eta_1 - \eta_2}{\eta_1 \eta_2}})}{(b-a)} -$$

$$\frac{\eta_1 \eta_2}{(\eta_1-1)(\eta_2-1)} \frac{(b^{\frac{\eta_1-1}{\eta_1}} - a^{\frac{\eta_1-1}{\eta_1}})(b^{\frac{\eta_2-1}{\eta_2}} - a^{\frac{\eta_2-1}{\eta_2}})}{(b-a)^2} \} \text{if } \eta_1 \neq 1,\ \eta_2 \neq 1,\ 1/\eta_1 + 1/\eta_2 \neq 1$$

$$= \frac{\Gamma(1+1/\eta_1)\,\Gamma((2\eta_1-1)/\eta_1)}{\lambda_1^{1/\eta_1}\;\lambda_2^{\eta_1/(\eta_1-1)}}\left[\frac{\log(b/a)}{(b-a)} - \frac{\eta_1^2}{(\eta_1-1)}\,\frac{\left(b^{\frac{\eta_1-1}{\eta_1}} - a^{\frac{\eta_1-1}{\eta_1}}\right)}{(b-a)^2}\left(b^{1/\eta_1} - a^{1/\eta_1}\right)\right]$$

if $1/\eta_1 + 1/\eta_1 = 1$

$$= \frac{\Gamma(1+1/\eta_i)}{\lambda_i^{1/\eta_i}\;\lambda_{i'}}\left[\frac{\eta_i\eta_{i'}\left(b^{\frac{\eta_i\eta_{i'}-\eta_i-\eta_{i'}}{\eta_i\eta_{i'}}} - a^{\frac{\eta_i\eta_{i'}-\eta_i-\eta_{i'}}{\eta_i\eta_{i'}}}\right)}{(\eta_i\eta_{i'}-\eta_i-\eta_{i'})(b-a)} - \frac{\eta_i}{(\eta_i-1)}\,\frac{\left(b^{\frac{\eta_i-1}{\eta_i}} - a^{\frac{\eta_i-1}{\eta_i}}\right)\log(b/a)}{(b-a)^2}\right.$$

if $\eta_i \neq 1,\ \eta_{i'} = 1$

$$= \frac{1}{(\lambda_1\ \lambda_2)}\left[\frac{1}{(ab)} - \frac{\log(b/a)^2}{(b-a)^2}\right] \quad \text{if } \eta_1 = \eta_2 = 1$$

For this model, the reliability function for a series system is

$$R_s(t) = \frac{[\exp(-b(\lambda_1 t^{\eta_1} + \lambda_2 t^{\eta_2})) - \exp[-a(\lambda_1 t^{\eta_1} + \lambda_2 t^{\eta_2})]]}{(b-a)(\lambda_1 t^{\eta_1} + \lambda_2 t^{\eta_2})} \tag{4.7}$$

and for a parallel system is

$$R_p(t) = \frac{[\exp(-b(\lambda_1 t^{\eta_1}) - \exp(-a\,\lambda_1 t^{\eta_1})}{(b-a)\,\lambda_1 t^{\eta_1}} + \frac{[\exp(-b\,\lambda_2 t^{\eta_2}) - \exp(-a\,\lambda_2 t^{\eta_2})}{(b-a)\,\lambda_2 t^{\eta_2}} - Rs(t) \tag{4.8}$$

Figure 4 shows the reliability for a series system and figure 5 for a parallel system under the uniform model for various combinations of a and b. Notice that when a = .25, b = .75, which corresponds to an operating environment which is less severe than the test environment, the system reliability is greater than that expected under independence, while when (a,b) = (1.25, 1.75) or (1, 2), which corresponds to an environment more severe than the test environment, the system reliability is smaller. Also when (a,b) contains 1, which corresponds to an environment which incurs the possibility of no differential effect from that found in the laboratory, there is little difference in the dependent and independent system reliability.

Acknowledgment

This work was supported by the U.S.Air Force Office of Scientific Research under contract AFOSR-82-0307.

References

Bagchi, S. B. and Samanta, K. C. (1984) : On Some Compound Bivariate Weibull Distributions In Life Testing. *IAPQR Transactions, Vol. 9*, 1, 41-48.

Bain, L. J. (1978). : *Statistical Analysis of Reliability and Life-testing Models*. Marcel Dekker, Inc., New York.

Basu, A. P. and Klein, J. P. (1982). : Some Recent Results in Competing Risk Theory. *Survival Analysis* Ed. by Crowley and Johnson The Institute of Mathematical Statistics, 216-227, Hayword.

Barlow, R. E. and Proschan, F. (1981) : *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart and Winston, New York

Clayton, D. G. (1978). : A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familiar Tendency in Chronic Disease Incidence. *Biometrika*, 65, 141-151.

Hougaard, P. (1985). : Multivariate Weibull and other Distributions for Dependent Failure Times., Novo Research Institute, Denmark

Hutchinson, T. P. (1981). : Compound Gamma Bivariate Distributions. *Metrika*, 28, 263-271.

Klein, J. P. and Moeschberger, M. L. (1984). : Consequences of Departure from Independence in Exponential Series Systems. *Technometrics*, 26 277-284.

Lee, S.(1986). :Inference for a Bivariate Survival Function Induced Through the Enviornment. Unpublished Ph.D. Thesis, Ohio State University.

Lehmann, E. L. (1959). : *Testing Statistical Hypotheses*. John Wiley and Sons, Inc., New York

Lindley D. V. and Singpurwalla, N. D. (1984). : Multivariate Distributions for the Reliability of a System of Components Sharing a Common Environment. George Washington University Technical Report.

Oakes, D. (1982). : A Model for Association in Bivariate Survival Data. *Journal of the Royal Statistical Society B*, 44, 414-422.

Schweizer, B. and Wolff, E. F. (1981). : On Nonparametric Measures of Dependence For Random Variables. *The Annals of Statistics*. 9, 879-885.

Takahasi, K. (1965). : Note on the multivariate Burr's Distribution. *Annals of the Institute of Statistical Mathematics, Tokyo*, 17, 257-260.

FIGURE 1

UPPER BOUND ON MAXIMAL CORRELATION FOR RANDOM ENVIORNMENT MODEL.

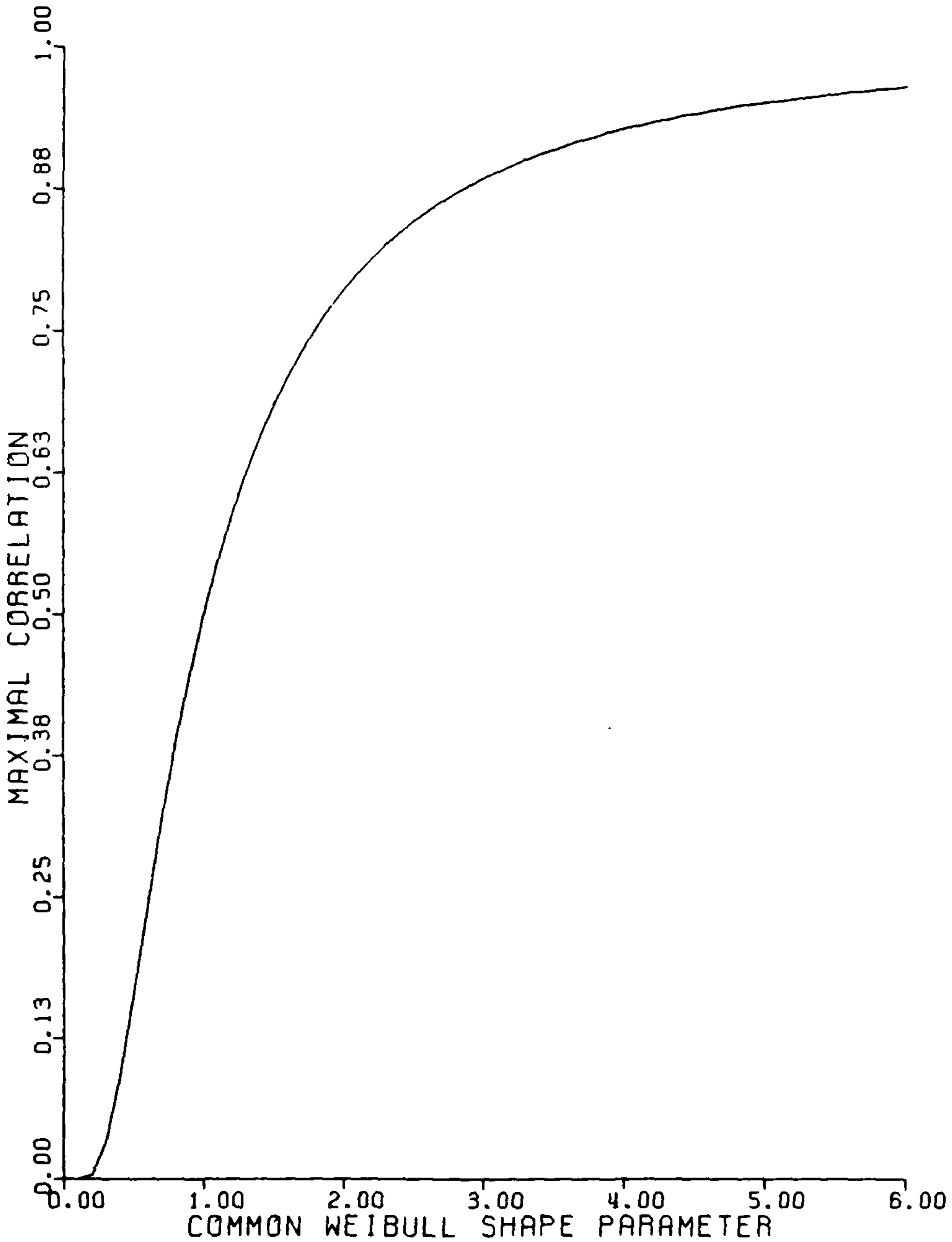


FIGURE 2

SERIES SYSTEM RELIABILITY UNDER GAMMA MODEL
FOR THE ENVIRONMENTAL STRESS.
$\lambda_1=1.0,\quad \lambda_2=2.0,\quad \eta_1=1.0,\quad \eta_2=1.0.$

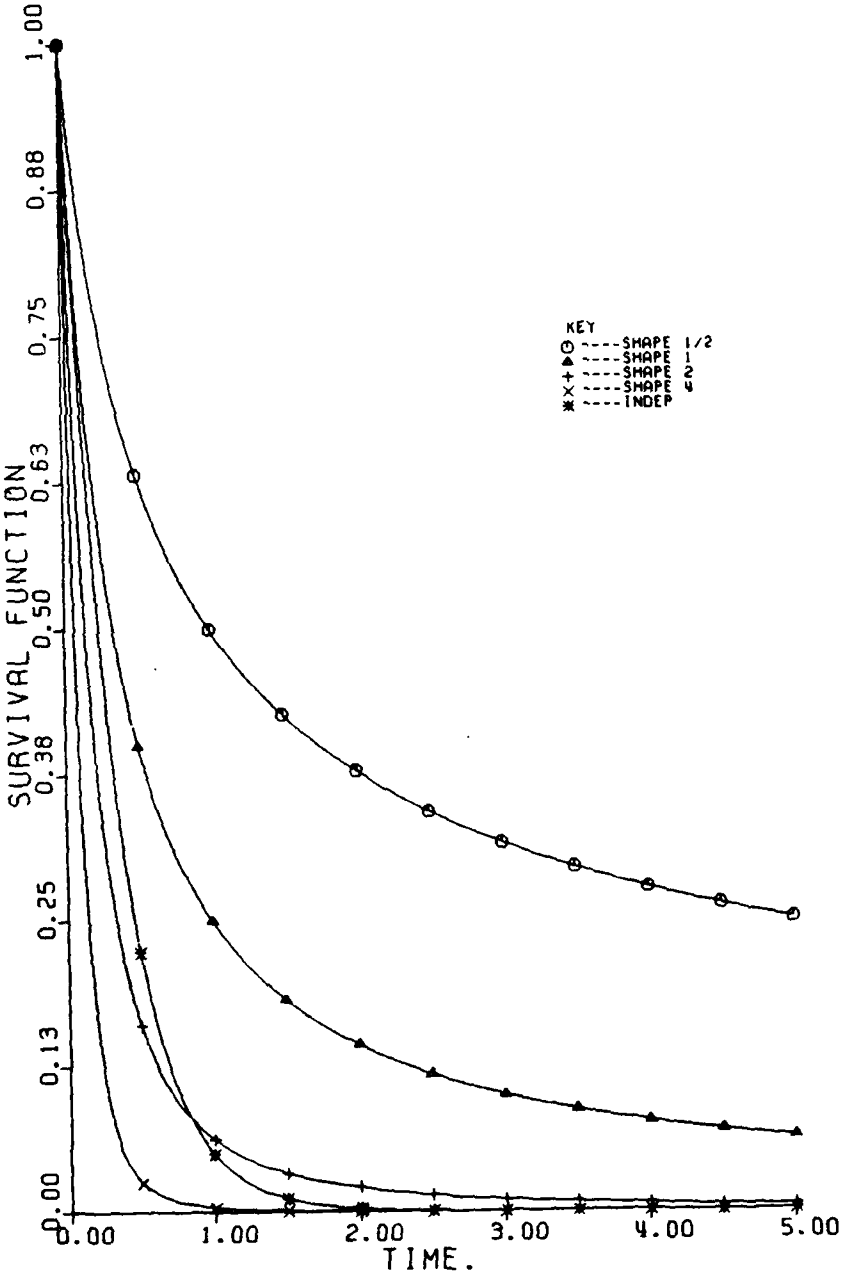


FIGURE 3

PARALLEL SYSTEM RELIABILITY UNDER GAMMA(A,1) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0$, $\lambda_2=2.0$, $\eta_1=1.0$, $\eta_2=1.0$.

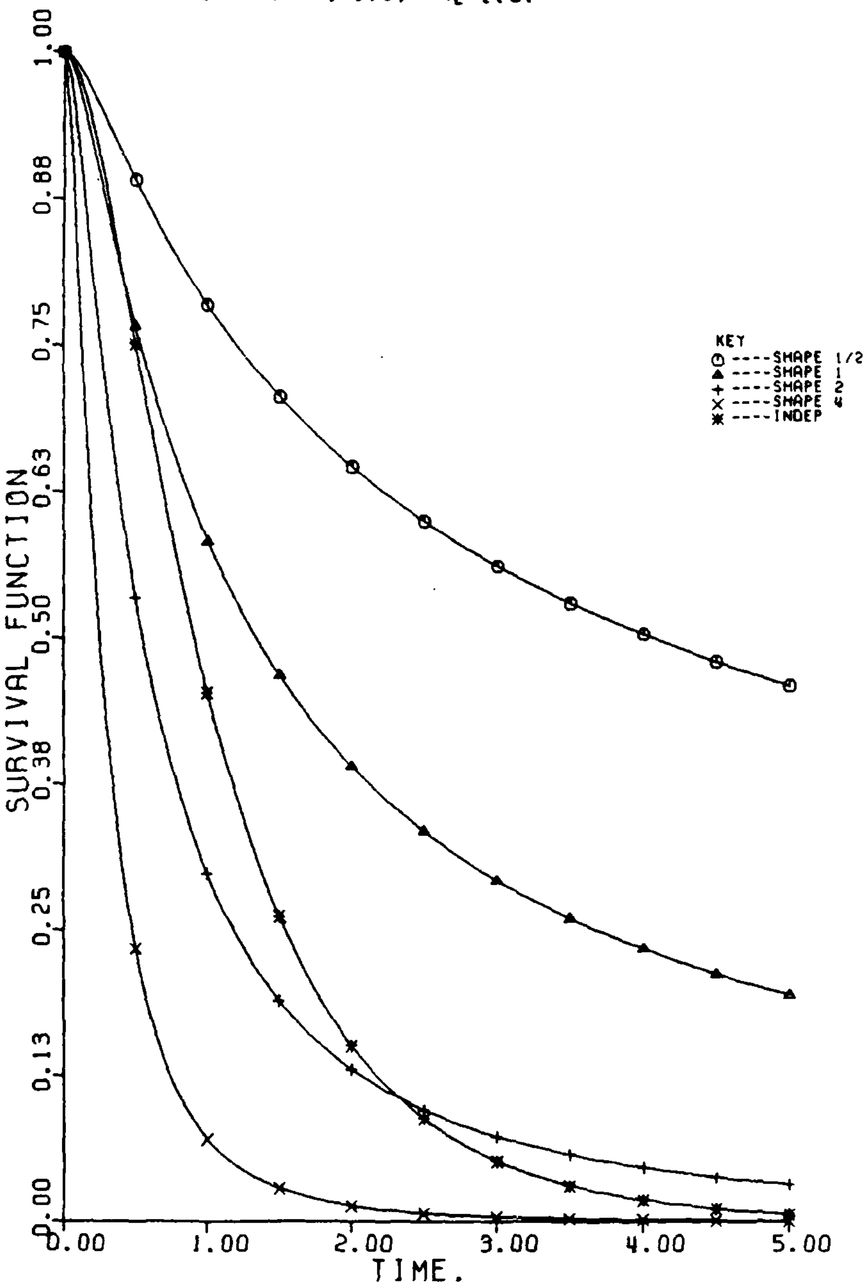


FIGURE 4

SERIES SYSTEM RELIABILITY UNDER UNIF(A,B) MODEL
FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0,\quad \lambda_2=2.0,\quad \eta_1=1.0,\quad \eta_2=1.0.$

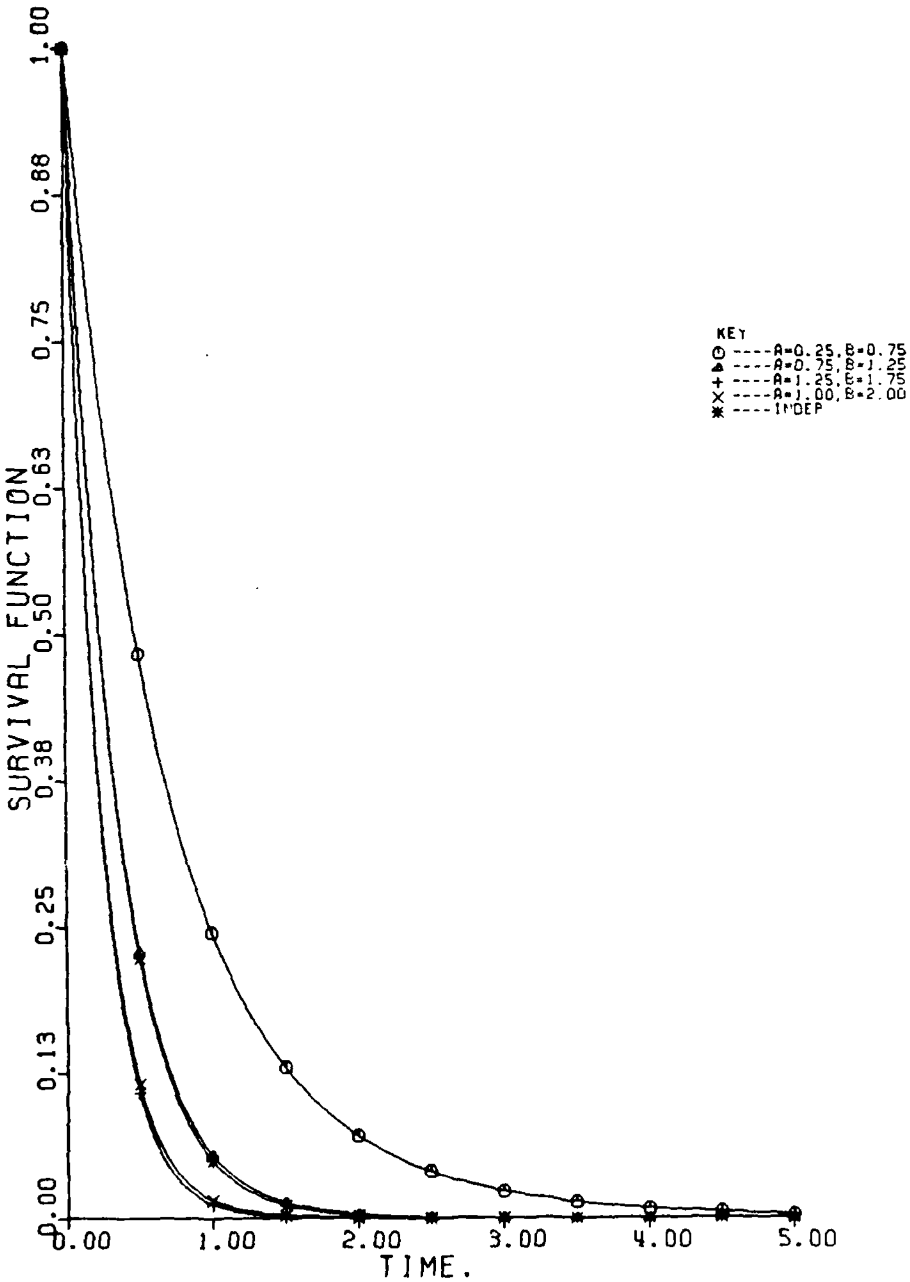
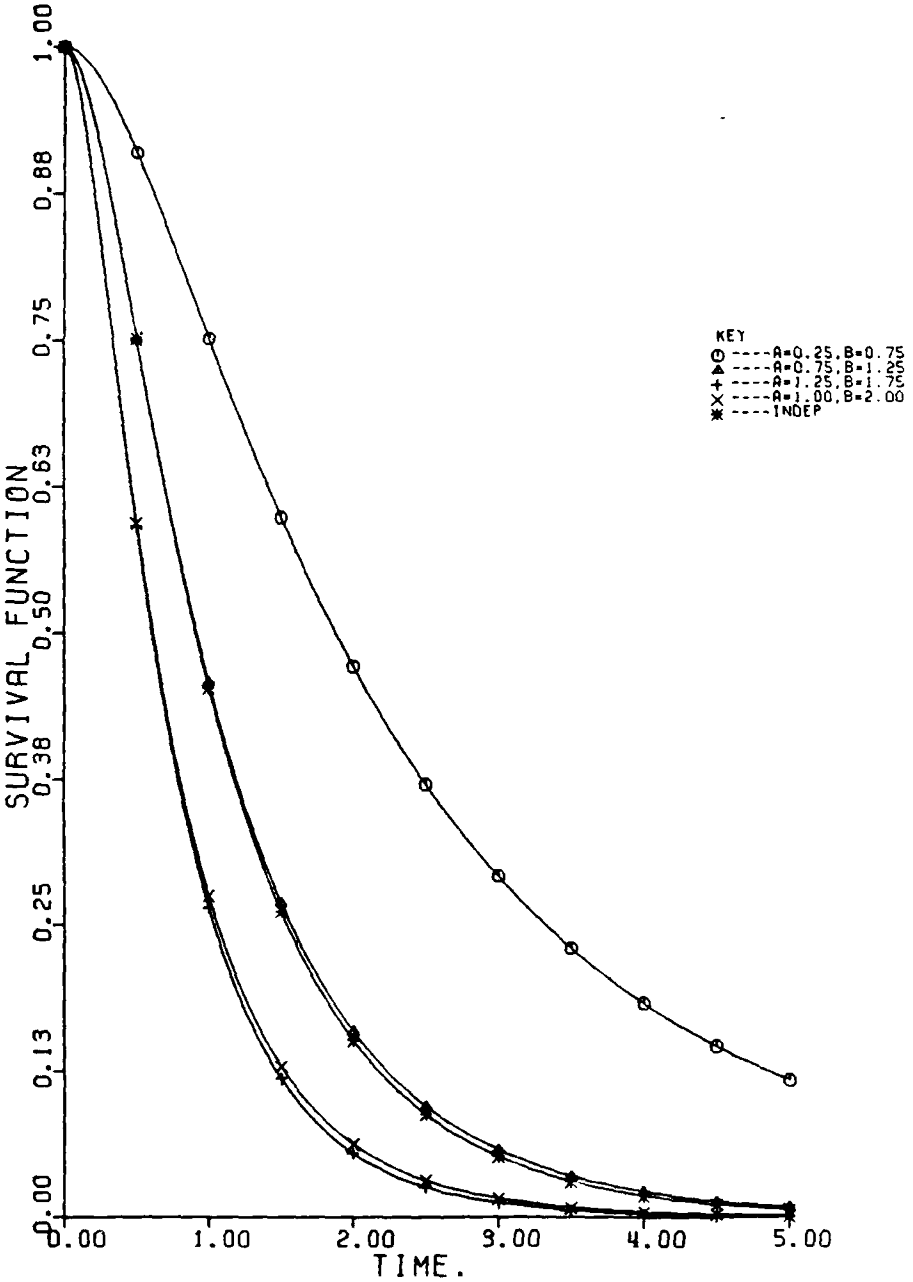


FIGURE 5

PARALLEL SYSTEM RELIABILITY UNDER UNIF(A,B) MODEL
FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0$, $\lambda_2=2.0$, $\eta_1=1.0$, $\eta_2=1.0$.



APPENDIX J

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY NA	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE NA	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION The Ohio State University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NM
6c. ADDRESS (City, State, and ZIP Code) 1314 Kinnear Road Columbus, Ohio 43212	7b. ADDRESS (City, State, and ZIP Code) AFOSR/NM Bldg 410 Bolling AFB DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-82-0307
8c. ADDRESS (City, State, and ZIP Code) AFOSR/NM Bldg 410 Bolling AFB DC 20332-6448	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61102F; PROJECT NO. 2304; TASK NO.; WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification) Statistical Method for Combining Laboratory and Field Data Based on a Random Environmental Stress Model

12. PERSONAL AUTHOR(S)
Sukhoon Lee and John P. Klein

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 9-1-82 TO 12-31-87	May 31, 1988	28

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP
XXXX	XXXXXXXXXX	XXX

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
System reliability, combination of experiments, scaled total time on test transform, maximum likelihood estimation

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

We present some techniques for analyzing combined experiments on systems tested under field conditions and components tested under controled conditions, under an assumed random environmental stress model for series system reliability. Based on this model we describe the maximum likelihood estimation of model parameters as well as other estimators based on the scaled total time on test statistic. The question of when it is advisable to perform diagnostic testing on the failed system to determine the failure modality in light of the escess cost of doing so is also studied. Lastly, a study of several tests for independence based on this model is made.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Maj. Brian Woodruff	22b. TELEPHONE (Include Area Code) (202) 767-5027 22c. OFFICE SYMBOL NM

STATISTICAL METHOD FOR COMBINING LABORATORY AND FIELD DATA BASED ON A RANDOM ENVIRONMENTAL STRESS MODEL

by

Sukhoon Lee & John P. Klein

February 1987

Technical Report No. 362

Sukhoon Lee
Department of Statistics
Chung-Nam National University
Dae-Jun, KOREA

John P. Klein
Department of Statistics
The Ohio State University
Columbus, Ohio USA

Statistical Method for Combining Laboratory and Field Data Based on a Random Environmental Stress Model

Sukhoon Lee
Department of Statistics
Chung-Nam National University
Dae-Jun, KOREA

John P. Klein
Department of Statistics
The Ohio State University
Columbus, Ohio USA

Summary

We present some techniques for analyzing combined experiments on systems tested under field conditions and components tested under controled conditions, under an assumed random enviornmental stress model for series system reliability. Based on this model we discribe the maximum likelihood estimation of model parameters as well as other estimators based on the scaled total time on test statistic. The question of when it is advisable to perform diagnostic testing on the failed systems to determine the failure modality in light of the excess cost of doing so is also studied. Lastly, a study of several tests for independence based on this model is made.

Key Words and Phrases: System Reliability, Combination of Experiments, Scaled Total Time on Test Transform, Maximum Likelihood Estimation.

1. Introduction

Consider a p - component series system where the failure of any one of the p components causes the system to fail. A common assumption made in analyzing such systems is that the potential (unobservable) times to failure of the components are statistically independent. Thus the assumption is not testable due to the identifiability problem. (See Basu and Klein (1982) for details). However, Klein and Moeschberger (1984) show that an investigator can be appreciably misled in modeling competing risks by erroneously assuming independence.

Recently several authors(Hutchinson(1981), Bagchi and Samanta(1984), Lindley and Singpurwalla(1984), Oakes(1982), Hoggard(1985), and Lee and Klein (1987)) have proposed models for dependent series systems based on a common random environmental factor which affects all components in the system under operating conditions. Inference for all the model parameters is not possible, based only on field data from the series systems, due to the identifiability dillema. However, in many situations, one has available additional data on the individual component reliabilities obtained in the laboratory during the testing and design process. Incorporating this data with the field data allows us to estimate model parameters and check a possible assumption of independence. In this paper we consider estimation of the model parameters for a two component series system based on a common environmental stress model. In section 2 we describe the particular model (Klein and Lee(1987)) on which the inference is made and the experimental setting. In section 3 maximum likelihood estimators are discussed and in section 4 we focus on estimating components' hazard rates. In section 5, a graphical estimation procedure is presented. In section 6 a comparision of the estimators obtained will be made through a small scale Monte Carlo study. Finally we comment on some tests for independence and present a small study.

2. Model

For simplicity we shall consider only the two component series system. The model for the system, as fully developed in Lee and Klein (1987) is as follows: Suppose that under ideal conditions, as encountered in the laboratory, the lifetimes of the two components are X_0 and Y_0 with reliability functions F_0 and G_0, respectively. The two components are linked into a system and put into the field. The effect of the operating environment is randomly improve or degrade each component by a factor Z with the distribution function H(z). That is, under field conditions the conditional reliability of the two components, given Z=z, are $F_0{}^z$ and $G_0{}^z$, respectively. Note that a value of $Z < 1$ implies a simultaneous improvement in component reliability, and $Z > 1$ implies a joint degradation. Further, we assume that conditional on Z the components are functioning independently.

In the sequel we assume that X_0 and Y_0 are exponential with hazard rates λ_1, λ_2 (i.e. $F_0(x) = \exp(-\lambda_1 x)$). This assumption is commonly made in reliability studies (c.f. Mann and Grubbs (1974), Chao (1981), Miyamura (1982), Boardman and Kendall (1970)). We also assume the random environmental factor, Z, follows a gamma distribution with density function is $h(z) = \{\Gamma(\alpha)\}^{-1} \beta^{\alpha} \exp(-z/\beta)\ z^{\alpha-1}$. This distribution is chosen for its analytical tractability and its flexibility in allowing for a variety of shapes including the exponential and bell shaped density. With these assumptions the reliability function for the system is

$$R(t) = E_Z \{ F_0{}^z(t)\ G_0{}^z(t)\} = (1+ (\lambda_1+\lambda_2)t/\beta)^{-\alpha}. \qquad (2.1).$$

This model has been proposed with a different parameterization by Lindley and Singpurwalla (1985) and Houggard (1985).

For this model, we note that the system reliability depends only on two parameters $\theta = (\lambda_1+\lambda_2)/\beta$ and α, so that if we had data only from systems on test in the operating environment, the only identifiable parameters are α and θ. However, in many instances we have extensive data on the performance of the components under controlled condition so that the whole experiment consists of three distinct parts.One experiment (EI) is done on the first component, A, under controlled condition, such as found in the laboratory and another independent experiment (EII) is performed on the second component, B, under controlled conditions. The third experiment (EIII) is carried out on the series systems under operating conditions.Sample data from the first two parts consist of times to failure of each component. The last part consists of the failure times of the system and an indicator variable which tells us which component causes the system to fail.

Let

$X_{0,i}$ = Lifetime of the i-th component A in EI, $i = 1, 2, \cdots, n$;

$Y_{0,j}$ = Lifetime of the j-th component B in EII, $j = 1,2,\cdots,m$;

n = Number of component A's put on the test under controlled conditions;

m = Number of component B's put on the test under controlled conditions;

$S_{0,1}=X_{0,1}+...X_{0,n}$; and $S_{0,2}=Y_{0,1}+...+Y_{0,m}$. Based on experiment EIII let

δ_i = The indicator variable whose value is equal to 1 if the A component failed first and otherwise equal to 0;

M $= \delta_1 + \delta_2 + \cdots + \delta_s$, the total number of system failures from component A;

T_i = Lifetime of the i-th system;

s = Number of the systems put on the test under operating conditions;

λ_1 = Hazard rate of the component A under controlled conditions;

λ_2 = Hazard rate of the component B under controlled conditions;

α = The shape parameter of the gamma distribution;

β = The scale parameter of the gamma distribution.

3. Maximum Likelihood Estimators

In this section we consider maximum likelihood estimation based on the three independent samples described in section 2. The relevant log likelihood is (up to constants)

$$\log L = n\cdot\log\lambda_1 + m\cdot\log\lambda_2 - \lambda_1 S_{o,1} - \lambda_2 S_{o,2} + M\log\lambda_1 + (s - M)\log\lambda_2$$

$$+ s\log\alpha - s\log\beta - (\alpha + 1)\sum_{k=1}^{s}\log\{1 + (\lambda_1+\lambda_2)\, T_k / \beta\}\,. \tag{3.1}$$

Taking derivatives we obtain the following system of four nonlinear equations:

i) $n / \lambda_1 - S_{o,1} + M / \lambda_1 - s / (\lambda_1 + \lambda_2) = 0$

ii) $m / \lambda_2 - S_{o,2} + (s - M) / \lambda_2 - s / (\lambda_1 + \lambda_2) = 0$

iii) $\dfrac{s}{\alpha} - \displaystyle\sum_{k=1}^{s} \log\{1 + (\lambda_1+\lambda_2)\, T_k / \beta\} = 0$

iv) $$\sum_{k=1}^{s} \frac{T_k}{1 + (\lambda_1+\lambda_2)\, T_k / \beta} - \frac{s\beta}{(\alpha+1)(\lambda_1+\lambda_2)} = 0\,. \tag{3.2}$$

The first two equations are to be solved for λ_1, λ_2 to obtain $\lambda_{11}, \lambda_{21}$. We may obtain α_{mle}, β_{mle} by solving the last two equations. In other words, the problem of finding the parameter values maximizing logL consists of the two parts: the first to find the values of λ_1, λ_2 maximizing $\log L_1$ and the second to find those values of α and $\theta = (\lambda_1 + \lambda_2) / \beta$ maximizing $\log L_2$, where

$$\log L_1 = n\cdot\log\lambda_1 + m\cdot\log\lambda_2 - \lambda_1 S_{o,1} - \lambda_2 S_{o,2} + \log({}_sC_M) + M\log\lambda_1 + (s - M)\log\lambda_2 \tag{3.3}$$

$$- s\log(\lambda_1+\lambda_2),$$ and

$$\log L_2 = s\log\alpha + s\log\theta - (\alpha + 1) \sum_{k=1}^{s} \log(1 + \theta T_k). \qquad (3.4)$$

It should be noted that the M.L.E's of λ_1, λ_2 depend on the samples of the components under controlled conditions and M, the number of systems which fail due to the failure of component A only, while the actual system lifetimes are used to estimate α and θ.

Returning to the estimation problem, the M.L.E.'s of λ_1, λ_2 are easily calculated as

$\lambda_{11} = \{-r_1 + (r_1^2 + 4r_o r_2)^{1/2}\} / 2r_2$ if $r_2 > 0$, and $\{-r_1 - (r_1^2 + 4r_o r_2)^{1/2}\} / 2r_2$ otherwise. The

other estimator λ_{21} is computed as $(n_C - S_{o1}\lambda_{11}) / S_{o2}$,

where $r_2 = S_{o,1} \cdot S_{o,2} - S_{o,1}^2$, $r_1 = (n_c + n_A) S_{o,1} + (s - n_A) S_{o,2}$, $r_o = n_C \cdot n_A$,

$n_C = n + m$; and $n_A = n + M$. Since λ_1 and λ_2 could be estimated based on EI or EII alone a natural question is " How much are these estimators improved by adding information from the system experiment, EIII?" Since this question is of independent interest we will discuss this problem in the next section.

Noting that $\log L_2$ is a function of α and θ, the two likelihood equations are

$$\frac{\partial \log L_2}{\partial \alpha} = \frac{s}{\alpha} - \sum_{k=1}^{s} \log\{1 + \theta T_k\} = 0\text{, and} \qquad (3.5)$$

$$\frac{\partial \log L_2}{\partial \theta} = \frac{s}{\theta} - (\alpha + 1) \sum_{k=1}^{s} \frac{T_k}{1 + \theta T_k} = 0. \qquad (3.6)$$

Solving 3.5 for α we obtain $\alpha_{mle} = s/\{\Sigma \log\{1 + \theta T_k)\}$. Substituting this value into 3.6 we obtain the following equation which is to be solved for θ

$$s / \theta - f_2(\theta) [\, s / f_1(\theta) + 1 \,] = 0 \quad \text{where} \quad f_1(\theta) = \Sigma \log(1 + \theta T_k)\text{, and} \qquad (3.7)$$

$$f_2(\theta) = \Sigma\, T_k / (1 + \theta T_k).$$

One can show that θ_{mle} is in the allowable parameter space if the observation T_k's satisfies

$$s \sum_{k=1}^{s} T_k^2 - 2 \left(\sum_{k=1} T_k \right)^2 > 0. \tag{3.8}$$

In the case that the data does not satisfy the condition (3.8) we would have a M.L.E. of θ at $\theta = 0$ which leads to ∞ as a M.L.E. of α. In such a case the reliability for the series system becomes

$$\lim_{\substack{\alpha \to \infty \\ \alpha/\beta \to \mu}} R_s(t) = \lim_{\substack{\alpha \to \infty \\ \alpha/\beta \to \mu}} (1 + (\lambda_1+\lambda_2)t/\beta)^{-\alpha} = \exp(-\mu\ (\lambda_1+\lambda_2)\ t) \tag{3.9}$$

so that we conclude that the series system has constant hazard rate and it seems reasonable to carry out the inference procedure accordingly. This condition is satisfied if the sample standard deviation is larger than the sample mean.

In addition, we note that the estimate of α can be less than one, which implies that the mean system reliability is infinite. To study the properties of these estimators and a graphical estimator which will be discribed later, a small scale Monte Carlo study will be presented in section 6.

4. A Note on the Estimation of the Components' Hazard Rates

As discussed in the previous section the M.L.E.'s of λ_1, λ_2, the components' hazard rates under controlled condition are obtained through the likelihood function $\log L_1$ which is constructed from three independent samples, one based on each component tested separately and one based on system data. However, the only contribution from the system data to the likelihood function for λ_1, λ_2 is the information as to which component has caused the system failure, while the contribution from the component data consists of their lifetimes. A natural question is how much does the field data improve the precision of our estimators of λ_1 and λ_2?

First, we compute the asymptotic variances of M.L.E.'s of λ_1, λ_2 obtained in the previous section and compare them with the variances of the M.L.E. computed only from the component samples. Second, we investigate some possible strategies for determining sample sizes under cost constraints which may occur if it costs to check which component caused the system to fail.

We assume that the sample sizes of both components are same, that is, $n = m = N$. Suppose that the ratio of the component sample size to the total sample size, $N/(2N + s)$, goes to c as both N and s go to ∞. Now the total information in the study is $I(\lambda)$ and

$$I(\lambda) = c\, I_1(\lambda) + c\, I_2(\lambda) + (1-2c)\, I_3(\lambda) \tag{4.1}$$

where $\lambda = (\lambda_1, \lambda_2)$ and $I_i(\lambda)$ is the information matrix based on the i-th experiment.

From the variance - covariance matrix $I^{-1}(\lambda)$ of M.L.E.s of (λ_1, λ_2)

$$\begin{matrix} \dfrac{\lambda_1^2(\lambda_1+\lambda_2)^2\{c(\lambda_1^2+\lambda_2^2)+\lambda_1\lambda_2\}}{c^2(\lambda_1^2-\lambda_2^2)^2+2c\lambda_1\lambda_2(\lambda_1+\lambda_2)^2} & \dfrac{\lambda_1^2\lambda_2^2(\lambda_1+\lambda_2)^2(1-2c)}{c^2(\lambda_1^2-\lambda_2^2)^2+2c\lambda_1\lambda_2(\lambda_1+\lambda_2)^2} \\ \\ \dfrac{\lambda_1^2\lambda_2^2(\lambda_1+\lambda_2)^2(1-2c)}{c^2(\lambda_1^2-\lambda_2^2)^2+2c\lambda_1\lambda_2(\lambda_1+\lambda_2)^2} & \dfrac{\lambda_2^2(\lambda_1+\lambda_2)^2\{c(\lambda_1^2+\lambda_2^2)+\lambda_1\lambda_2\}}{c^2(\lambda_1^2-\lambda_2^2)^2+2c\lambda_1\lambda_2(\lambda_1+\lambda_2)^2} \end{matrix}, \tag{4.2}$$

we obtain, by Theorem 6.1 of Lehmann (1983), $(2N+s)^{1/2}(\lambda_{11}-\lambda_1) \longrightarrow N(0, I_{11}^{-1})$, $(2N+s)^{1/2}(\lambda_{21}-\lambda_2) \longrightarrow N(0, I_{22}^{-1})$, where I_{ij}^{-1} is the i-jth entry in the matrix $I^{-1}(\lambda)$.

Based on the laboratory experiment EI only, the maximum likelihood estimator of λ_1 is $\lambda_{10} = N/S_{0,1}$ and $\sqrt{N}(\lambda_{10}-\lambda_1) \longrightarrow N(0, \lambda_1^2)$. Thus the asymptotic relative efficiency (A.R.E.) of λ_{10} to λ_{11} is $e_1 = V(\lambda_{11})/V(\lambda_{10})$ and

$$e_1 = c\frac{I_{11}^{-1}}{\lambda_1^2} = \frac{(\lambda_1+\lambda_2)^2\lambda_1\lambda_2 + c(\lambda_1+\lambda_2)^2(\lambda_1^2+\lambda_2^2)}{2(\lambda_1+\lambda_2)^2\lambda_1\lambda_2 + c(\lambda_1+\lambda_2)^2(\lambda_1-\lambda_2)^2}. \tag{4.3}$$

If $\lambda_2 = k\lambda_1$, then $e_1 = \{k + c(k^2+1)\} / \{2k + c(k-1)^2\}$. $\quad$ (4.4)

Figure 4.1 shows the A.R.E. of λ_{11} to λ_{10} at varying ratios of component sample size to total sample size for five different k values. Notice that for fixed c the more similar the hazard rates of the two components are, the smaller the A.R.E. is. That is, the more information the system sample contributes to reducing the variance. Figure 4.2 shows the A.R.E. at varying k's for given ratio of sample sizes.Note that c=0 corresponds to the situation where N is very small compared to s, c=1/6 to the case where s is of the order 4 times N, c=1/3 to the case where s is almost same as N, c=2/5 to the case where N of the order 1/2 times N and c=1/2 to the case where s is very small compared to N. The above result is also valid for the estimators, λ_{21} and λ_{20} since all the formulae are symmetric in λ_1 and λ_2.

In the previous section we saw that estimating the scale parameter β of the gamma distribution involves estimating $\lambda_1 + \lambda_2$ rather than λ_1 or λ_2 themselves so that we shall turn to comparison of the variances of $\lambda_{11} + \lambda_{21}$ and $\lambda_{10} + \lambda_{20}$. Using the same procedure as before we are led to the A.R.E. of $\lambda_{10} + \lambda_{20}$ to $\lambda_{11} + \lambda_{21}$ which is $e_2=V(\lambda_{11}+\lambda_{12})/V(\lambda_{10}+\lambda_{20})$ and

$$e_2 = \frac{(\lambda_1 + \lambda_2)^2 \ \{ \lambda_1\lambda_2 + c \ (\lambda_1 - \lambda_2)^2 \}}{(\lambda_1^2 + \lambda_2^2) \ \{2\,\lambda_1\lambda_2 + c \ (\lambda_1 - \lambda_2)^2 \}} . \tag{4.5}$$

Setting $\lambda_1 = k\,\lambda_2$,

$$e_2 = \left(1 + \frac{2k}{(k^2 + 1)} \right) \left(1 - \frac{k}{c\,(k-1)^2 + 2k} \right) . \tag{4.6}$$

The plots of the A.R.E.'s are shown for various k's and c's in figure 4.3 and 4.4.Here it is found that when the hazard rates of two components are identical or very similar the information from the system does not contribute to a reduction of the asymptotic variance. An intuitive explanation for this is that when the two hazard rates are similar the information from the system, which only contributes information on which of the two hazard rates is larger, through the numbers of systems which fail from each type of component failure, contributes least to the inference on the magnitude of the sum of the two hazard rates.

There are several problems which may be considered in the light of the above results. We will discuss two of these:

i) If we suppose that it costs to check the cause of system failure, when is it resonable to do so.

ii) If we are allowed to randomly check only a fraction of the systems, how many systems should be checked to achieve minimum variance under some constraints?

To investigate the above two problems we assume that the sample size of the system sample is fixed at s and that the sample sizes and the unit price of testing both components are the same. Let P_T be the total remaining allowable cost after administrative costs and the costs of collecting system life data are removed and, let P_U be the cost of testing a component in EI and EII, and let P_c be the additional cost of checking a system to determine its failure mode. Suppose these costs P_T , P_U, and P_c are predetermined.The sample sizes are assumed to be reasonably large so that asymptotic results hold.

We consider question (i). We see that

$N_c = P_T / 2P_U$: The maximum number of each component we could test when systems are not checked.

$R = P_c / P_U$: The ratio of the costs,

$Q = N_c / s$: The ratio of the component sample size if systems are not checked to the system sample size, and,

$N = N_c - R \cdot s / 2$: if we check all systems.

Our goal is to find the maximum value of R such that $V(\lambda_{11,N}+\lambda_{21,N}) < V(\lambda_{10,N_c}+\lambda_{20,N_c})$ where the subscripts N, N_c denote the component sample sizes when the estimators are computed. Noting that $V(\lambda_{11,N}+\lambda_{21,N})$ is approximated by

$$\frac{1}{2N+s} \quad \frac{(\lambda_1+\lambda_2)^2}{c_f} \quad \left[\frac{c_f(\lambda_1-\lambda_2)^2+\lambda_1\lambda_2}{c_f(\lambda_1-\lambda_2)^2+2\lambda_1\lambda_2}\right] \text{ where } c_f = N/(2N+s), \text{ and}$$

$V(\lambda_{10,N_c}+\lambda_{20,N_c})$ by $(\lambda_1^2+\lambda_2^2)/N_c$, we obtain the following approximate ratio

$$\frac{V(\lambda_{11,N}+\lambda_{21,N})}{V(\lambda_{10,N_c}+\lambda_{20,N_c})} = \frac{Q}{Q - 0.5R}\left[\frac{(k+1)^2}{k^2+1}\left(\frac{\frac{Q-0.5R}{2Q+(1-R)}(k-1)^2+k}{\frac{Q-0.5R}{2Q+(1-R)}(k-1)^2+2k}\right)\right]. \tag{4.7}$$

If we let r(R, Q, k) denote the above ratio of variances, after some algebraic manipulation we see that $r(R,Q,k) < 1$ implies

$$R^2(k+1)^2 - R(2Qk^2+4Qk+4k+2Q) + \{(k-1)^2/(k^2+1)\}4Qk > 0. \tag{4.8}$$

Thus the maximum value of R for which checking of the system failure mode is advisable is

$$\frac{1}{(k+1)^2}\left(B - \{B^2 - 4Qk(k-1)^2\}^{1/2}\right), \text{ where } B = Qk^2 + 2Qk + 2k + Q. \tag{4.9}$$

Figure 4.5 shows the maximum value of R for each Q at different k's. For example, suppose that Q is equal to 10 computed using the predetemined values s, P_T, and P_U.If we also assume that the ratio of the hazard rates is ,3 , which might be guessed through past experience then this figure tells that if the relative cost, R, is less than 0.1 it is advisable to check the systems. As discussed before, if the two components have the same hazard rates then it is not recommended to do so.

Considering the question (ii), we find the number of systems to be checked to achieve optimal results in terms of the variance of the sum of the two M.L.E.'s under the given constraints. Let s^* be the number of systems to be checked among the s systems, and set $y = s^* / N_c$. Noting that $N=N_c - 0.5\,R\,s^*$, and recalling that the ratio of component sample to complete sample size c is $N/(2N+s^*)$, we obtain the asymptotic variance of $V(\lambda_{11,N}+\lambda_{21,N})$ ignoring the constant terms as

$$q(y, R, k) = \frac{1}{1 - 0.5Ry}\left[\frac{\{c(k-1)^2+k\}}{\{c(k-1)^2+2k\}}\right]. \tag{4.10}$$

The derivative of q with respect to y is,

$$\frac{dq}{dy} = \frac{q_1(y)}{(1 - 0.5Ry)^2 [(k+1)^2 + y (2k - kR - 0.5k^2R - 0.5R)]^2} \tag{4.11}$$

where
$$\begin{aligned} q_1(y) = {} & y^2 [0.5R (k^2 + 1) - k] [0.5R (k+1)^2 - 2k] (R/2) \\ & + y [R (k^2 + 1) (2k - 0.5R (k+1)^2)] \\ & + [0.5(k+1)^2(k^2+1)] R - k(k-1)^2 . \end{aligned}$$

Studying the function $q_1(y)$ in detail we find that $q(y, R, k)$ is minimized at $y = 0$. That is, no contribution is made by checking the system failure mode if the relative cost P_c / P_U is larger than $[2k(k-1)^2] / [(k+1)^2(k^2+1)]$, and that if the relative cost is smaller than the above ratio the optimal number of systems to be checked is,

$$N_c \left[\frac{\sqrt{2k(k-1)}}{(k - 0.5R(k^2+1))\sqrt{R(2k - 0.5R(k+1)^2)}} - \frac{(k^2+1)}{k - 0.5R(k^2+1)} \right]. \tag{4.12}$$

Figure 4.6 shows the optimal fraction $y = s^*/N_c$ at the allowable R's for k = 2, 3, and 5. For example, suppose we have the idea that the ratio of the hazard rates, k is equal to 5, and that the relative cost R is equal to .1. Then this figure tells that the optimal number of the systems to be checked is 1.5 times as much as N_c.

6. Graphical Inference

In this section, we discuss a graphical approach which is helpful in visualizing the condition of existence of the estimators and also the degree of dependence as well as in checking feasibility of the model. Later in this section we suggest estimators based on this graphical approach. Throughout this section we assume the same model as in the section 2. However we shall handle the model as if the component hazard rates λ_1 and λ_2 were known, based on data from the laboratory experiment, since the estimation of λ_1 and λ_2 presents little difficulty and has been discussed in detail in section 4. Thus, the model in this section is that the lifetime, T, of a system in the operating environment has a survival function $R_s(t) = (1 + \theta t)^{-\alpha}$. (5.1)

The method we present in this section is based on the scaled total time on test (STTOT) plot of Barlow and Campo (1975). They have presented a graphical approach to failure data analysis for arbitrary distributions, using the total time on test transforms introduced and discussed in Barlow et. al. (1972).

Suppose A(t) is the cummulative distribution function of T.

The STTOT transform for T is defined by

$$ST_A(t) = \int_0^{A^{-1}(t)} R_S(u)\,du \Big/ \int_0^{A^{-1}(1)} R_S(u)\,du$$

$$= 1 - (1-t)^{(\alpha-1)/\alpha} \quad \text{for } \alpha > 1. \tag{5.2}$$

Here we note that $ST_A(t)$ depends only on the shape parameter α. Figure 5.1 shows the form of the STTOT transform for several values of α. Notice that for all α, the STTOT transform is below the 45 $^{\circ}$ line (which corresponds to exponential system life) since the hazard rate of the series system is decreasing.This figure tells us that the smaller the shape parameter is, the more dependence that is induced.

The emperical total time on test at $T_{s,r}$ is defined by $V_{s,r} = T_{s,1}+T_{s,2}+\cdots+(s-r+1)T_{s,r}$ where $T_{s,1}, T_{s,2}, \cdots, T_{s,s}$ are the ordered system failure times, and plotting $V_{s,r}/V_{s,s}$ versus r/s for r = 1, 2, ⋯,s, we obtain so called the emprical STTOT plot. Since $V_{s,r}/V_{s,s}$ converges to $ST_A(t)$ with probability one and uniformly in $0 \le t \le 1$ as $s \longrightarrow \infty$ and $r/s \longrightarrow t$, the STTOT plot can be compared to the figure 5.1 for a graphical check of the model's validity. We can also use the STTOT transform to obtain estimators of the shape parameter α Let $C_i = \log(1-i/s)$ and $D_i = \log(1- V_{s,i} / V_{s,s})$, i = 1, ..., s-1. From $ST_A(t) = 1\ (1 - t)^{(\alpha-1)/\alpha}$ we have $\log(1 - ST_A(t)) = (1 - 1/\alpha)\log(1 - t)$ so that $D_i = (1-1/\alpha)C_i$, $i = 1, ..., s-1$.

First we consider , as a reasonable estimator of α , the value of α which minimizes the squared distances between D_i and $(1 - 1/\alpha)\, C_i$. That is,

$$\sum_{i=1}^{s-1} (D_i - (1-1/\alpha)\, C_i)^2 .$$

The resulting estimator is $\alpha_{ls} = \Sigma C_i^2 / (\Sigma C_i^2 - \Sigma C_i D_i)$ (5.3)

which is in the parameter space if $\Sigma C_i^2 > \Sigma C_i D_i$. A better estimator should be obtained by weighting the D_i's differently since for $i < j$, $\mathrm{Var}(D_i) < \mathrm{Var}(D_j)$. The variance of D_i depends on the unknown parameter α so we weight by the variance of D_i computed under an assumed exponential distribution. If $T_1, T_2, \cdots, T_s$ are assumed to follow an exponential distribution, then $[1 - V_{s,r}/V_{s,s}]$ follows a beta distribution with parameters s-r and r for r = 1,2, ...,s-1.Noting that the r-th order statistics of a sample of size s-1 from a uniform distribution follows a beta distribution with parameters r and s-r one can show that $-D_i$ is the i-th order statistics of a sample of size s-1 from a standard exponential distribution. Hence the variance of D_i in that case is

$$V_i = \sum_{j=1}^{i} \frac{1}{(s-j)^2}, \quad i = 1, \ldots, s-1$$

so that a weighted least squares estimator of α is

$$\alpha_{wls} = \Sigma C_i^2/V_i \, / \, \left(\Sigma \frac{C_i^2}{V_i} - \Sigma \frac{C_i D_i}{V_i}\right) \quad \text{if } \Sigma C_i^2/V_i > \Sigma C_i D_i/V_i. \tag{5.4}$$

Once we have obtained an estimator of α by either of the two least squares estimators, we substitute this value into (3.4) and solve this equation numerically for θ_{ls} or θ_{wls}. We note that the unique root of the equation lies between $1/(\alpha T_{s,s})$ and $1/(\alpha T_{s,1})$.

Due to the computational complexity of these estimators, analytic properties of these estimators are not available, so a small scale Monte Carlo study was performed in the next section to compare these estimators with three other estimators, the M.L.E. in section 3, the Method of Moments Estimator (M.M.E), and one suggested by Hui and Berger (1983).

The M.M.E.s, found by equating the first two sample and theoretical moments, so that

$$\alpha_{mme} = 1 + \frac{sE_2}{s\,E_2 - 2\,E_1^2} \qquad \text{if } s\,E_2 > 2\,E_1^2 \text{ , and} \tag{5.5}$$

$$\theta_{mme} = \frac{s\,E_2 - 2\,E_1^2}{E_1 \cdot E_2}\text{ ,} \qquad \text{where } E_1 = \sum_{k=1}^{s} T_k\text{ ,} \qquad E_2 = \sum_{k=1}^{s} T_k^2 \tag{5.6}$$

Hui and Berger (1983) have suggested estimators of α and θ in a different context. To avoid difficulties of maximizing the loglikelihood function with two unknown parameters they have suggested a modified method of moments estimator as follows., The estimator of α is the solution to

$$-\sum_{i=1}^{s} \log\left(1 + \frac{s}{E_1\,\alpha} T_i\right) + \frac{s(\alpha+1)}{E_1\,\alpha^2} \sum_{i=1}^{s} \frac{T_j}{1 + sT_i/(E_1 \cdot \alpha)} = 0, \tag{5.7}$$

and the estimator of θ is $[\,\alpha\,E_1 / s\,]^{-1}$. It is possible that there is no finite solution to this equation. With an argument similar to that used in M.L.E. case it can be shown that a sufficient condition for a finite solution to (5.7) is that $s\,E_2 > 2\,E_1^2$,which is the same condition needed for exsistance of the M.L.E. and M.M.E..

6. Monte Carlo Study

In this section we compare the estimators of the shape parameter α and the scale parameter θ through a small scale Monte Carlo study. The main comparisions of interest are done in terms of the bias, standard deviation of the estimates of α and θ, and the number of samples where the estimators exist. Also the estimators of system reliabilty at $t_o = 0.1006$ are compared. Random samples of size s = 15, 30, 50, 75, or 100 were generated with $\lambda_1 + \lambda_2 = 3$, $\beta = 3$, so $\theta = 1$ and $\alpha = 2, 3, 5$. 1000 samples were generated for each combination of s and α. The bias, standard deviation of the estimates and NS, the number of samples where the estimator exists is reported in table 6.1 for α, table 6.2 for θ, and in table 6.3 for an estimator of the system reliability obtained from (5.1) at t_o . The true system reliability at t_o is .8255 when $\alpha = 2$, .75 when $\alpha =$

Table 6.1: Bias and Standard Deviation (SD) of Estimators of α

Sample Size	Estimator	α = 2 NS	α = 2 Bias	α = 2 SD	α = 3 NS	α = 3 Bias	α = 3 SD	α = 5 NS	α = 5 Bias	α = 5 SD
15	mle	769	4.5	29.	642	7.3	39.	522	38.7	573.
	wls	852	4.8	41.	753	36.4	843.	665	8.2	53.
	*	766	1.3	5.	636	1.3	9.	516	-0.7	5.
	ls	762	6.8	49.	653	13.1	149.	558	30.3	493.
	mme	770	9.1	37.	643	16.8	77.	522	69.8	925.
	ber	770	14.4	65.	643	26.0	114.	522	112.9	1505.
30	mle	916	2.8	20.	809	5.7	30.	674	20.4	148.
	wls	953	4.7	37.	870	13.8	141.	752	9.3	68.
	*	912	1.1	3.	804	1.8	6.	660	3.2	29.
	ls	877	6.4	52.	768	11.9	100.	669	13.7	109.
	mme	916	6.1	32.	809	9.9	104.	674	31.7	202.
	ber	916	10.0	52.	809	17.7	68.	674	56.3	347.
50	mle	979	5.8	114	916	3.6	18.	801	7.6	39.
	wls	981	1.7	10.	935	6.9	65.	850	9.0	97.
	*	976	1.0	3.	912	2.9	29.	787	2.0	10.
	ls	956	4.0	16.	864	6.4	33.	756	13.4	88.
	mme	979	8.5	131.	916	6.6	25.	801	11.4	52.
	ber	979	15.4	241.	916	12.5	42.	801	23.4	91.
75	mle	996	0.9	4.	963	2.5	14.	893	12.8	139.
	wls	998	1.0	5.	977	2.8	17.	915	8.0	94.
	*	996	1.0	5.	958	1.3	5.	878	2.9	16.
	ls	974	2.4	12.	925	11.6	144.	823	6.6	22.
	mme	996	2.2	4.	963	4.7	24.	893	15.3	122.
	ber	996	4.5	8.	963	9.6	38.	893	32.6	260
100	mle	999	0.5	3.	978	1.7	7.	892	9.5	84.
	wls	1000	1.7	35.	989	2.1	12.	913	19.6	307.
	*	999	0.6	2.	978	1.3	5.	879	13.7	273.
	ls	989	1.5	9.	956	3.7	19.	835	11.0	81.
	mme	999	1.7	5.	978	3.0	9.	892	13.0	120
	ber	999	3.7	8.	978	7.2	15.	892	27.1	203.

'*' represents the weighted least squares estimator restricted to those samples where all estimators exist.

Table 6.2: Bias and Standard Deviation (SD) of θ

Sample Size	Estimator	α = 2 NS	Bias	SD	α = 3 NS	Bias	SD	α = 5 NS	Bias	SD
15	mle	769	0.356	1.702	642	0.691	1.900	522	1.352	3.109
	wls	852	-.102	0.742	753	0.210	1.049	665	0.715	1.609
	*	766	-.027	0.025	636	0.390	1.040	516	1.084	1.599
	ls	782	-.192	0.729	653	0.119	1.031	558	0.578	1.536
	mme	770	-.683	0.205	643	-.513	0.348	522	-.238	0.601
	ber	770	-.803	0.122	643	-.705	0.203	522	-.546	0.348
30	mle	916	0.112	0.919	809	0.175	1.049	674	0.558	1.609
	wls	953	-.135	0.580	870	0.000	0.757	752	0.366	1.118
	*	912	-.100	0.567	804	0.074	0.740	660	0.523	1.104
	ls	877	-.254	0.586	769	-.096	0.745	669	0.180	1.026
	mme	916	-.623	0.192	809	-.469	0.338	674	-.199	0.576
	ber	916	-.798	0.095	809	-.725	0.160	674	-.584	0.282
50	mle	979	0.016	0.648	916	0.075	0.766	801	0.256	1.559
	wls	989	-.126	0.492	935	-.012	0.618	850	0.184	0.869
	*	976	-.115	0.486	912	0.011	0.609	787	0.267	0.850
	ls	956	-.263	0.514	864	-.112	0.663	756	0.105	0.863
	mme	979	-.575	0.184	916	-.404	0.333	801	-.193	0.541
	ber	979	-.792	0.079	916	-.718	0.135	801	-.615	0.232
75	mle	996	-.025	0.522	963	0.030	0.624	893	0.128	0.817
	wls	998	-.125	0.432	977	-.027	0.555	915	0.112	0.728
	*	996	-.124	0.431	958	-.010	0.546	878	0.153	0.715
	ls	974	-.247	0.275	925	-.144	0.603	827	0.014	0.747
	mme	996	-.541	0.174	963	-.375	0.322	893	-.189	0.535
	ber	996	-.790	0.065	963	-.717	0.120	893	-.628	0.210
100	mle	999	-.019	0.437	978	-.028	0.515	892	0.033	0.683
	wls	1000	-.101	0.381	989	-.075	0.472	913	0.020	0.628
	*	999	-.100	0.401	978	-.055	0.465	879	0.055	0.615
	ls	989	-.216	0.423	956	-.165	0.511	835	-.064	0.666
	mme	999	-.508	0.153	978	-.345	0.297	892	-.206	0.494
	ber	999	-.785	0.052	978	-.716	0.104	892	-.644	0.185

'*' represents the weighted least squares estimator restricted to those samples where all estimators exist.

Table 6.3: Bias and Standard Deviation (SD) of Estimators of System Reliability at $t_o = .1006$

Sample Size	Esti-mator	$\alpha = 2$ NS	Bias	SD	$\alpha = 3$ NS	Bias	SD	$\alpha = 5$ NS	Bias	SD
15	mle	769	-.012	.0647	642	-.018	.0815	522	-.029	.1011
	wls	852	-.004	.0586	753	-.010	.0767	665	-.024	.0967
	*	766	-.006	.0577	636	-.015	.0764	516	-.030	.0968
	ls	762	0.002	.0588	653	-.006	.0748	558	-.020	.0977
	mme	770	0.037	.0503	643	0.031	.0661	522	0.012	.0926
		769	0.064	.0463	643	0.067	.0616	522	0.054	.0921
30	mle	916	-.005	.0473	809	-.007	.0577	674	-.022	.0691
	wls	953	0.002	.0424	870	-.007	.0552	752	-.020	.0674
	*	912	0.001	.0426	804	-.006	.0544	660	-.027	.0665
	ls	877	0.010	.0434	769	0.002	.0551	669	-.013	.0651
	mme	916	0.037	.0357	809	0.024	.0490	674	0.001	.0613
	ber	916	0.069	.0335	809	0.062	.0472	674	0.045	.0608
50	mle	979	-.001	.0372	916	-.003	.0429	801	-.006	.0545
	wls	989	0.003	.0349	935	-.001	.0412	850	-.006	.0530
	*	976	0.003	.0348	912	-.002	.0411	787	-.008	.0528
	ls	956	0.010	.0359	864	0.005	.0431	756	-.002	.0532
	mme	979	0.035	.0300	916	0.022	.0366	801	0.010	.0494
	ber	979	0.071	.0233	916	0.066	.0340	801	0.055	.0485
75	mle	996	0.000	.0290	963	-.001	.0372	893	-.005	.0442
	wls	998	0.004	.0274	977	0.000	.0356	915	-.005	.0436
	*	996	0.003	.0274	958	0.000	.0357	878	-.006	.0430
	ls	974	0.010	.0292	925	0.007	.0374	827	-.002	.0431
	mme	996	0.034	.0244	963	0.021	.0327	893	0.007	.0406
	ber	996	0.072	.0238	963	0.067	.0313	893	0.051	.0380
100	mle	999	0.001	.0243	978	0.001	.0309	892	-.002	.0375
	wls	1000	0.004	.0234	984	0.002	.0299	913	-.001	.0392
	*	999	0.004	.0233	978	0.002	.0299	879	-.002	.0390
	ls	989	0.010	.0248	956	0.008	.0311	835	0.003	.0380
	mme	999	0.034	.0223	978	0.019	.0267	892	0.008	.0353
	ber	999	0.075	.0216	978	0.067	.0261	892	0.052	.0345

'*' represents the weighted least squares estimator restricted to those samples where all estimators exist.

3,and .619 when $\alpha = 5$. Also reported in each table is the bias and standard deviation of the weighted least square estimators when they are restricted to those samples where the other estimators exist.

From these tables we note that Berger's modified estimator performs very poorly. Also the weighted least squares estimator allows for estimation of parameters in many more samples when s is small. In general the maximum likelihood estimator out performs the other estimators, however, when the weighted least squares estimator is restricted to those samples where the maximum likelihood estimator exists, this estimator performs much better when s is small. The somewhat better performance of the M.L.E in terms of bias is deceptive since some of the estimates of α are less than one, which implies that the mean system reliability is infinite. Also the weighted least squares estimator of system reliability seems to out perform the other estimators of the system reliability in spite of its relatively poor performance as an estimator of θ. Our recommendation is to use the weighted least squares estimator since it more often provides estimators of the relevant parameters and is somewhat easier to compute.

7. Comment On Test for Dependence

In this section we discuss the problem of determining whether there is a dependence structure induced by an environmental factor. In our setting, we observe only the system failure times T_i with the assumption that the survival function of T_i is $R_s(t) = (1+\theta t)^{-\alpha}$. As pointed out in section 6, the graphical presentation indicates that the shape parameter, α only affects the dependence structure. Accordingly we will call the quantity $\gamma = 1/\alpha$ a measure of dependence induced by the environmental factor. Since it is resonable to assume finite mean system lifetime, that is, α is assumed to be greater than 1, γ varies from 0 to 1. If there is no dependence induced γ is equal to 0 and the closer γ is to 1 the more dependence induced

One possible statistics is constructed from the weighted least square estimator as

$$Q_s = (\Sigma\, C_i D_i / V_i) / (\Sigma\, C_i^2 / V_i) \qquad (7.1)$$

Under the null hypothesis of independence, $-D_i$ follows the distribution of the i-th order statistics among the sample of size s-1 from an exponential distribution so that Q_s is just a linear combination of exponential order statistics . Hence Q_s has the same distribution as a linear combination, $Q_s(z)$, of identically independent exponential random variables since the i-th exponential order statistics can be expressed as a linear function of s-1 independent standard exponentials. Correspondingly we have

$$Q_s(z) = \sum_{i=1}^{s-1} p_i Z_i \tag{7.2}$$

where Z_i is a random variable following the standard exponential distribution,

$$\text{and } p_i = \frac{1}{s-i} \sum_{j=i}^{s-1} \frac{C_j / V_j}{\sum_{k=1}^{s-1} C_k^2 / V_k} . \tag{7.3}$$

The exact distribution of $Q_s(z)$ is found in David (1981) as the mixture of exponentials,

$$f_{Q_s(z)}(t) = \sum_{i=1}^{s-1} \frac{w_i}{p_i} \exp\left(\frac{-t}{p_i}\right), \quad \text{where } w_i = p_i^{s-2} / \{\prod_{h \neq i} (p_h - p_i)\} . \tag{7.4}$$

On the other hand we can note that if γ goes to 1, Q_s tends to have smaller value. Table 7.1 shows the critical values of the standardized Q_s for different sample sizes with the type one error

probability $\alpha = .01, .05, .1$. Since the distribution under alternatives is hard to obtain a simulation study has been constructed to study the tests power which is discussed later.

A second test is based on the cummlative total time on test statistic, which has been introduced by Barlow et. al.(1972). From the cummulative total time on test statistics (CTTS) which they have defined as

$B_s = \sum_{r=1}^{s} \frac{V_{s,r}}{V_{s,s}}$ where $V_{s,r}$ is the total time on test defined in section 5, we develope a

test statistics R_s as

$$(12\,s)^{1/2}\left[\,s^{-1}\sum_{r=1}^{s}\frac{V_{s,r}}{V_{s,s}}-\frac{1}{2}\,\right], \tag{7.5}$$

which has smaller values as γ goes to 1.

Table 7.1: Critical Values of the StandardizedStatistics of Q_s

Sample Size	1 %	5 %	10 %
s = 15	-1.8880	-1.4540	-1.1968
s = 20	-1.9382	-1.4815	-1.2106
s = 25	-1.9796	-1.5000	-1.2194
s = 30	-2.0010	-1.5140	-1.2260
s = 35	-2.0334	-1.5243	-1.2307
s = 40	- 2.0526	-1.5322	-1.2345
s = 50	-2.0816	-1.5453	-1.2404

A third test statistic was introduced by Klefsjo (1983). He has used the property of convexity of the STTOT, $ST_A(t)$ and obtained the statistic,

$$K_s = \sum_{j=1}^{s} a_j \,\frac{(s-j+1)\,(T_{s,j}-T_{s,j-1})}{V_{s,s}} \tag{7.6}$$

where $a_j = \dfrac{\{(s-1)^3 j - 3\,(s+1)^2 j^2 + 2\,(s+1)\,j^3\}}{6}$, which has smaller values as γ goes to 1.

Using the asymptotic properties of linear combination of order statistics he has shown that under the null hypothesis, the test statistic K_s is asymptotically normally distributed. In addition to that, he has constructed a list of critical values from the exact distribution of K_s under the null hypothesis.

We have, by simulation, estimated the powers for the various shape parameter values. Sample sizes 20 and 50 have been studied.In figure 7.1 and figure 7.2 the estimated power curves for the three tests mentioned above are obtained by the following scheme. The total number of replication for each investigated γ-value, measure of dependence, which increases from .00 to .75 is 1000.

The significance levels are equal to .05. The three powers at each γ-value have been estimated from the same set of data. Our investigation leads us to the conclusion that Q_S and R_S outperform K_S and that the test statistics Q_S, which has been developed only for this specific model, is not better than R_S which can detect more general alternatives than Q_S. However, the test statistics Q_S, is simple to obtain while we are making a graphical inference on the shape parameter and is guaranteed to keep power as high as R_S.

ACKNOWLEGEMENT

This work was supported by the U.S.Air Force Office of Scientific Research under contract AFOSR-82-0307.

BIBLIOGRAPHY

Bagchi, S. B. and Samanta, K. C. (1984). On Some Compound Bivariate Weibull Distributions in Life Testing. IAPQR Transactions, Vol. 9, 1, 41-48.

Barlow, R. E., Bartholomew, D. J. , Bremner, J. M., and Brunk, H. D. (1972) Statistical Inference Under Order Restrictions. New York: John Wiley and Sons, Inc.

Barlow, R. E. and Campo, R. (1975). Total Time on Test Processes and Applications to Failure Data Analysis. Reliability and Fault Tree Analysis. (Barlow, Fussell, and Singpurwalla, eds.). Philadelphia: SIAM, 451-481.

Basu, A. P. and Klein, J. P. (1982). Some Recent Results in Competing Risk Theory. Survival Analysis (Crowley and Johnson, eds.). Hayword: The Institute of Mathematical Statistics, 216-227.

Boardman, T. J., and Kendall, P. J. (1970). Estimation in Compound Exponential Failure Models. Technometrics, 12, 891 -900.

Chao, A. (1981) Approximate Mean Squared Errors of Estimators of Reliability for k-out-of-m Systems in the Independent Exponential Case. Journal of the American Statistical Association, 76, 720-724.

David, H.A. (1981) Order Statistics. New York: John Wiley and Sons, Inc.

Hougaard, P. (1985). Multivariate Weibull and other Distributions for Dependent Failure Times. Novo Research Institute, Denmark.

Hui, S. L. and Berger, J. (1983). Empirical Bayes Estimation of Rates. Journal of the American Statistical Association 78, 753-760.

Hutchinson, T. P. (1981). Compound Gamma Bivariate Distributions. Metrika 28, 263-271.

Klein, J. P. and Moeschberger, M. L. (1984). Consequences of Departure from Independence in Exponential Series Systems. Technometrics 26, 277-284.

Klefsjo, B.(1983). Some Tests Against Aging Based on The Total Time on Test Transform. Communication Statistics A. Theory and Method 12, 907-927.

Lee, S . and Klein, J. P. (1987). Bivariate Models With Random Environmental Factor. IAPQR Transactions(to appear).

Lehmann, E. L. (1983). Theory of Point Estimation. New York: John Wiley and Sons, Inc.

Lindley D. V. and Singpurwalla, N. D. (1985). Multivariate Distributions for the Reliability of a System of Components Sharing a Common Environment. George Washington University Technical Report.

Mann, N. R. and Grubbs, F. E. (1974). Approximately Optimum Confidence Bounds for System Reliability Based on Component Test Data. Technometrics, 16, 335-347

Miyamura, T (1982). Estimating Component Failure Rates from Combined Component and Systems Data: Exponentially Distributed Component Lifetimes. Technometrics 24, 313-318.

Oakes, D. (1982). A Model for Association in Bivariate Survival Data. Journal of the Royal Statistical Society B, 44, 414-422.

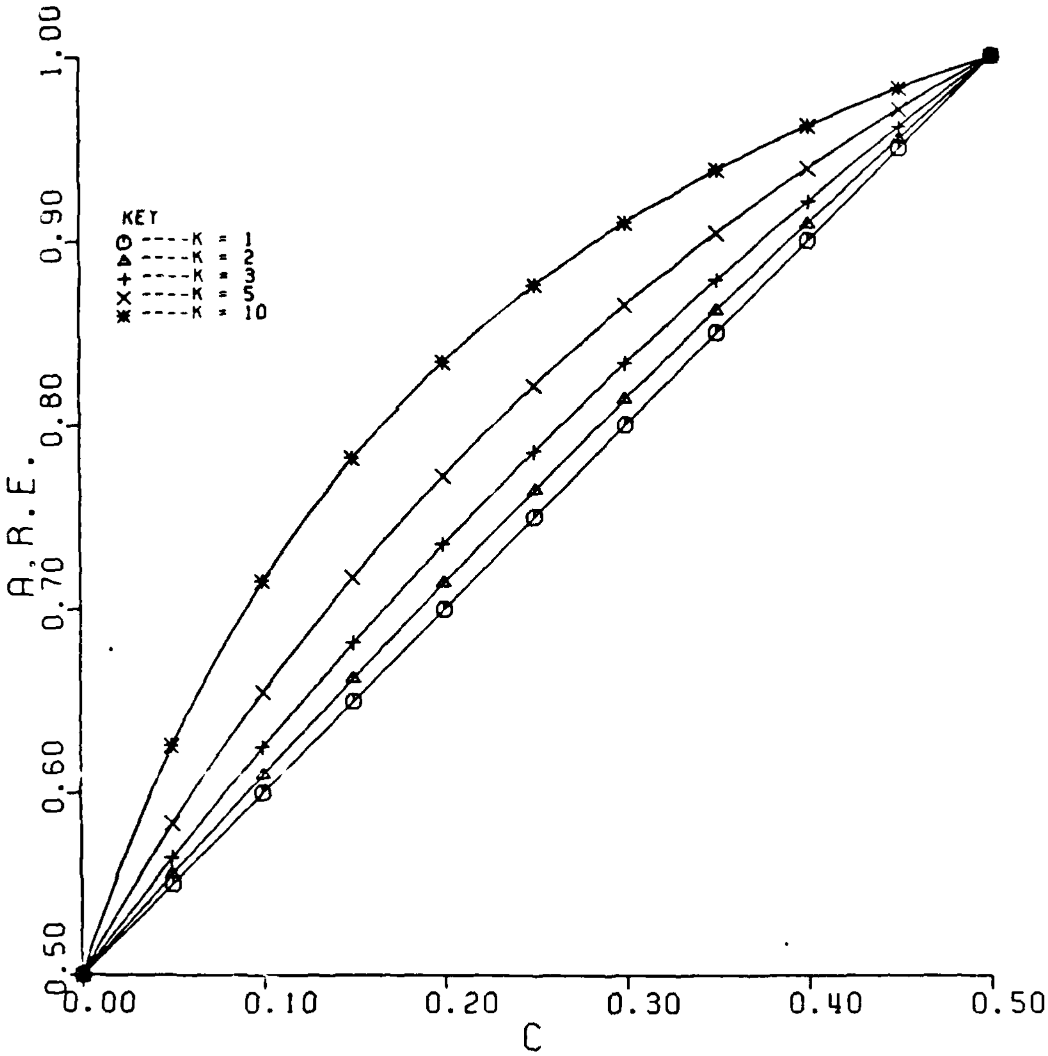


FIGURE 4.1

A.R.E. OF λ_{11} TO λ_{10} AS A FUNCTION OF C

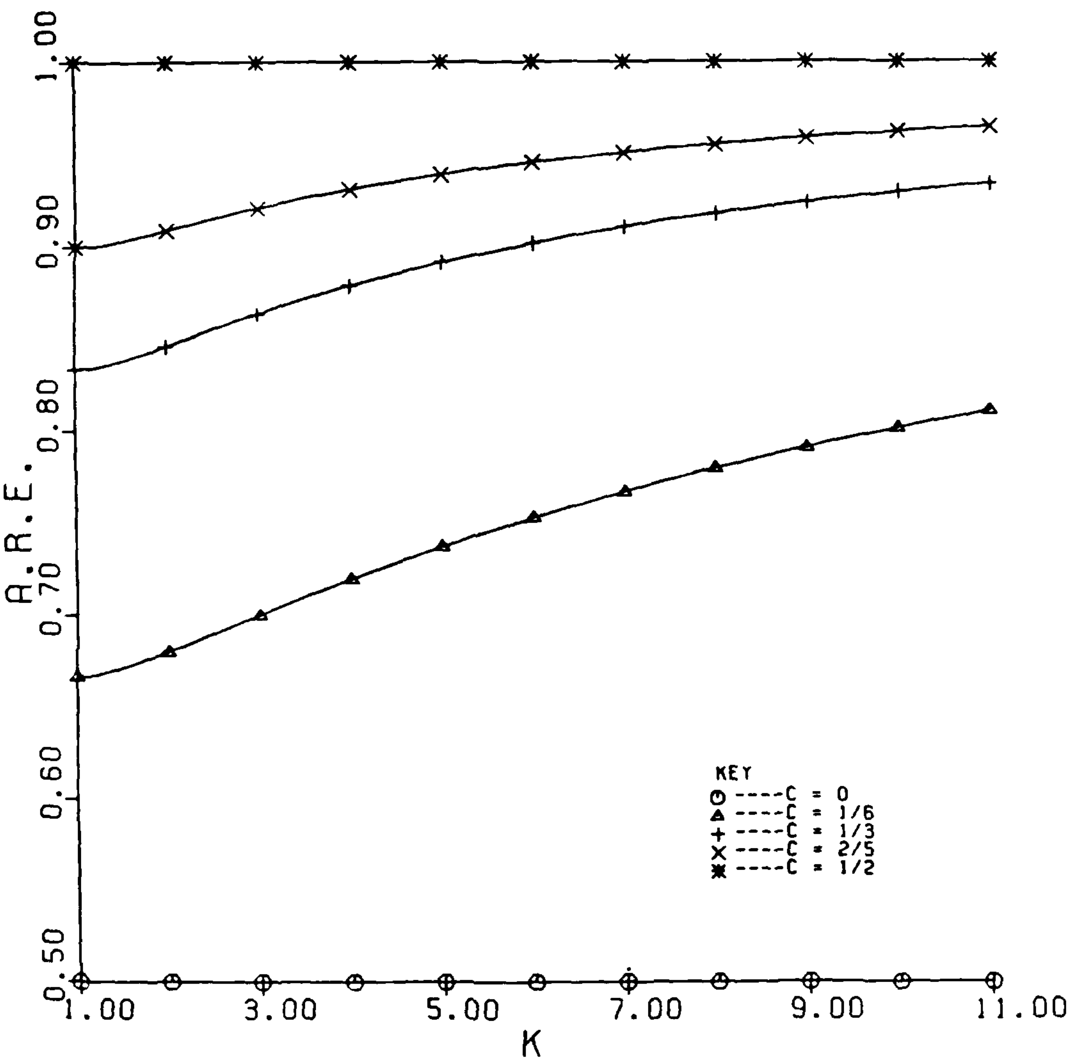


FIGURE 4.2

A.R.E. OF λ_{11} TO λ_{10} AS A FUNCTION OF K

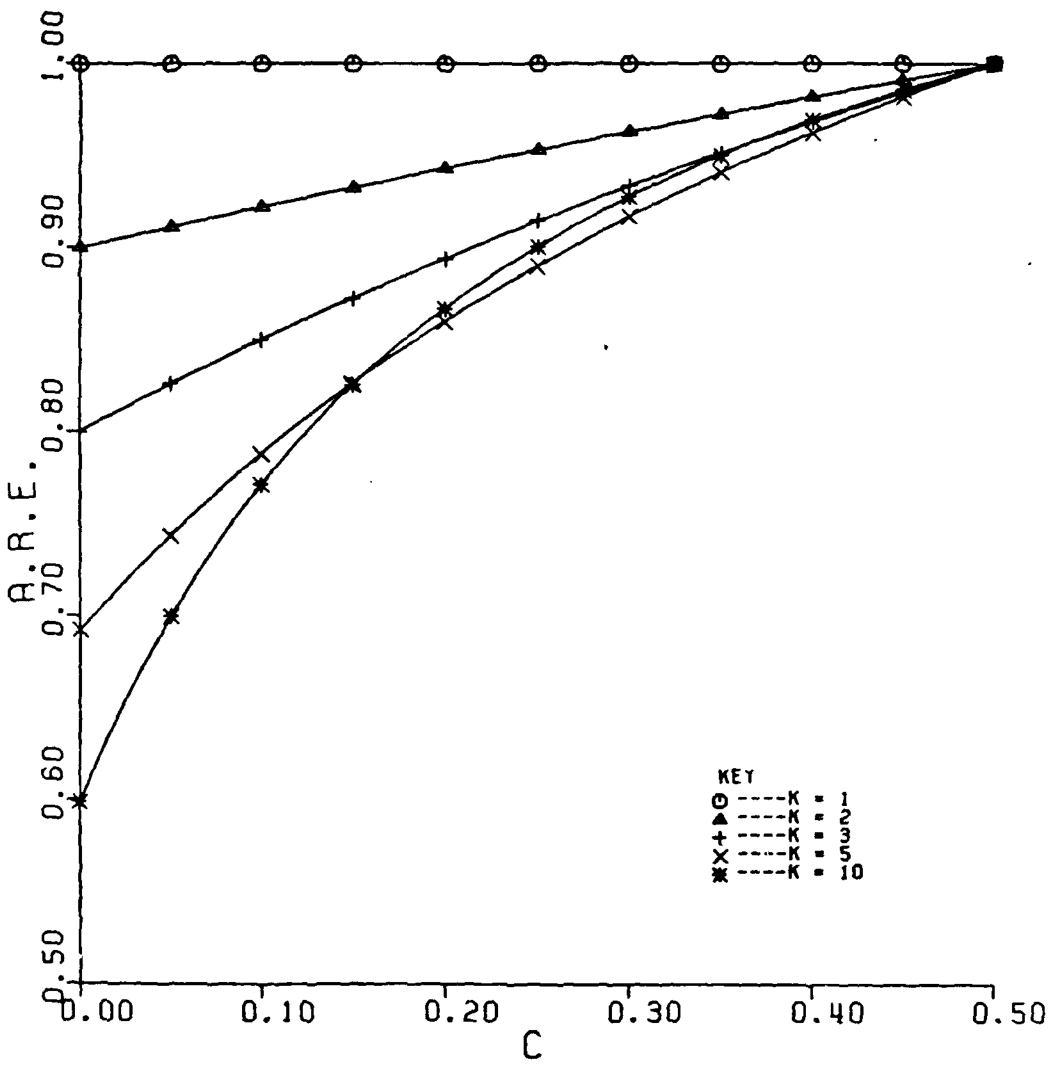


FIGURE 4.3

A.R.E. OF $\lambda_{11} + \lambda_{21}$ TO $\lambda_{10} + \lambda_{20}$ AS A FUNCTION OF C.

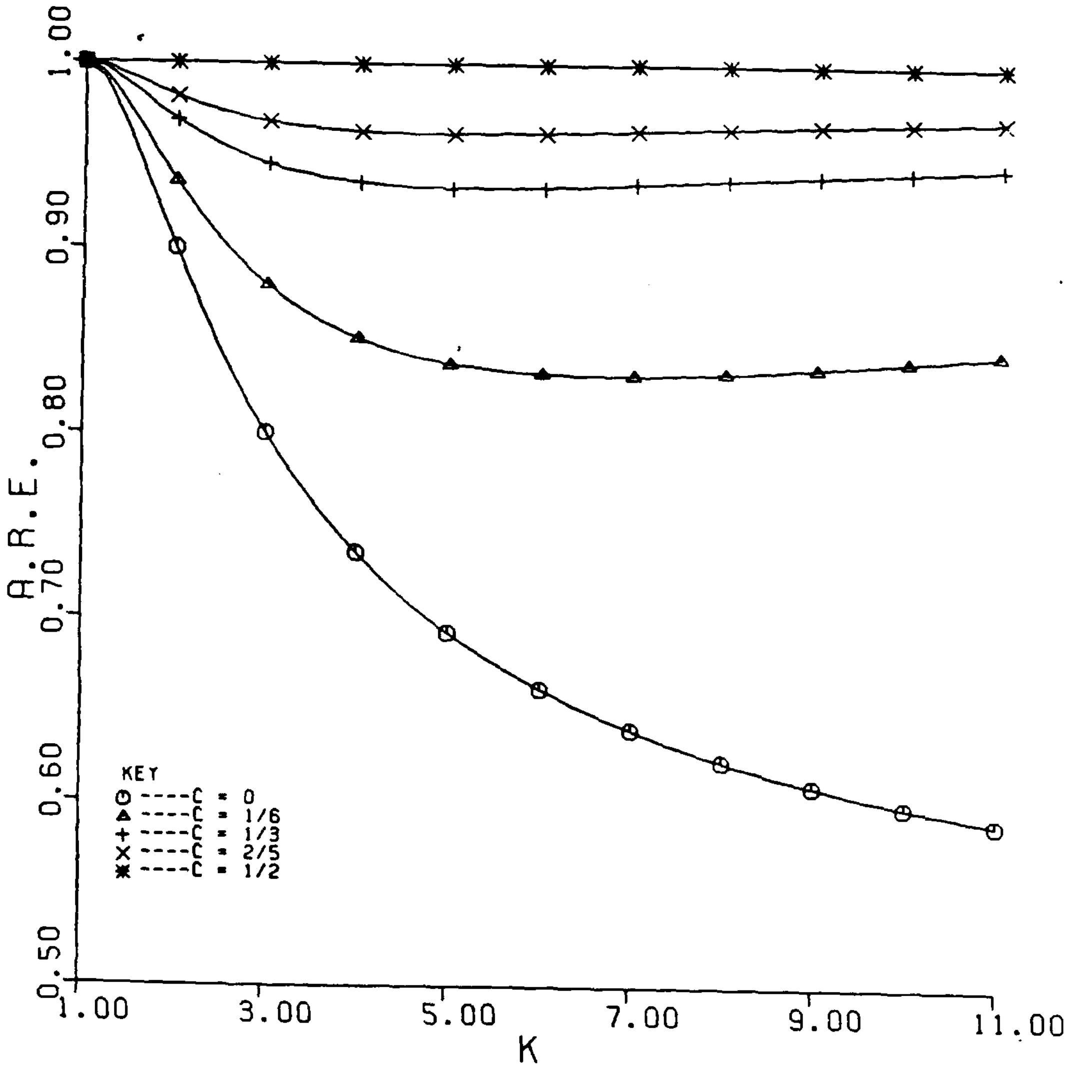


FIGURE 4.4

A.R.E. OF $\lambda_{11} + \lambda_{21}$ TO $\lambda_{10} + \lambda_{20}$ AS A FUNCTION OF K.

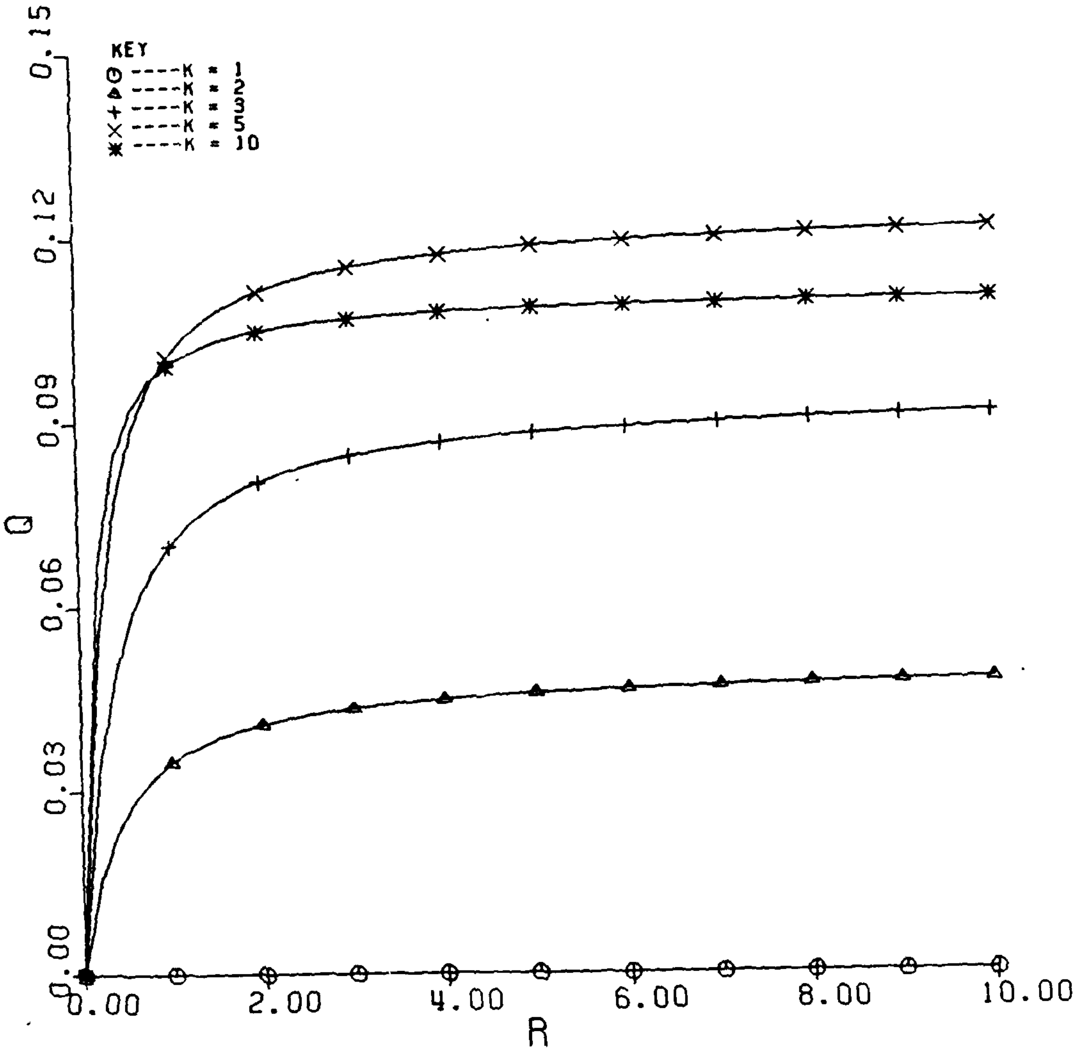


FIGURE 4.5

MAXIMUM RELATIVE COST, R, AT WHICH SYSTEM FAILURE MODE INFORMATION IS WORTH COLLECTING.

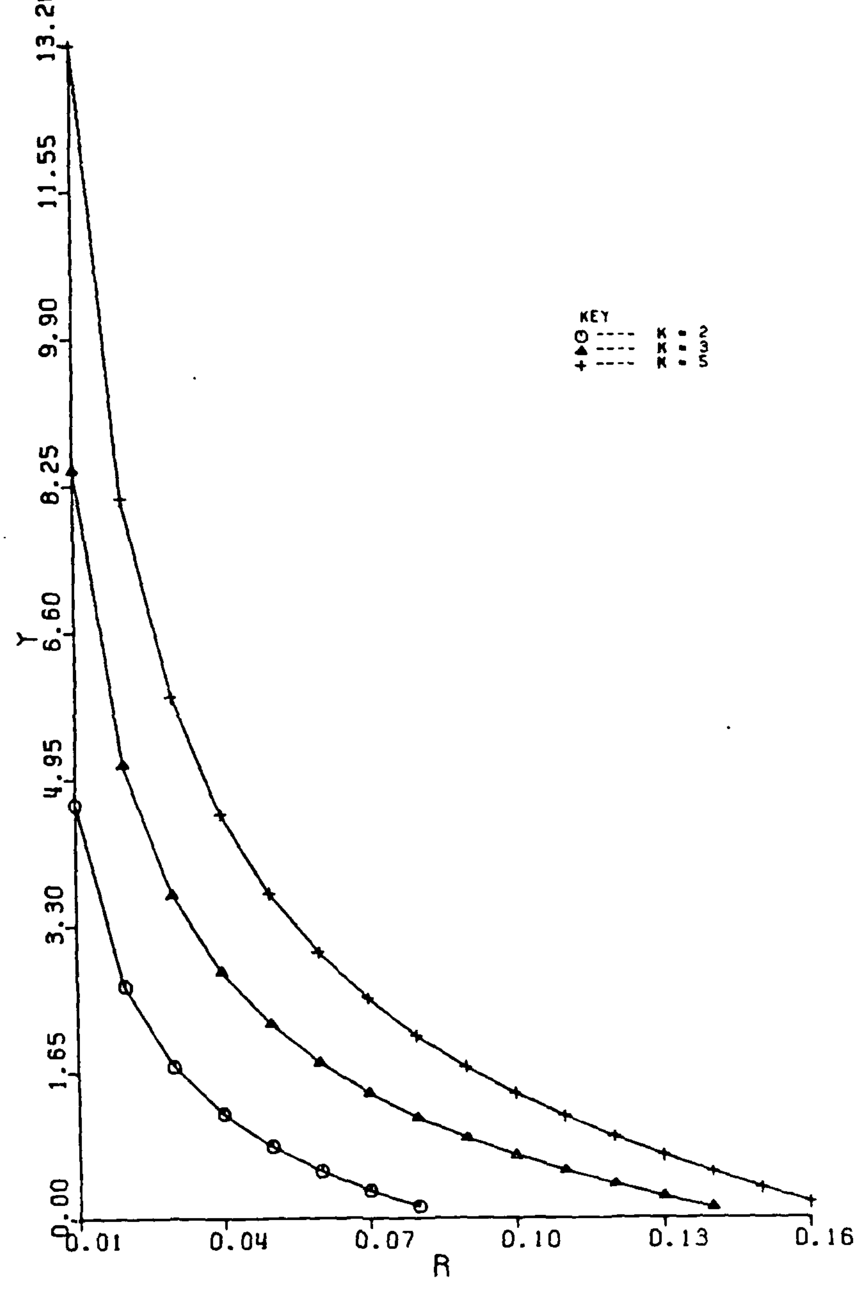


FIGURE 4.6

OPTIMAL FRACTION Y, OF THE SYSTEMS TO MAKE
FAILURE MODE DETERMINATIONS ON FOR A FIXED COST

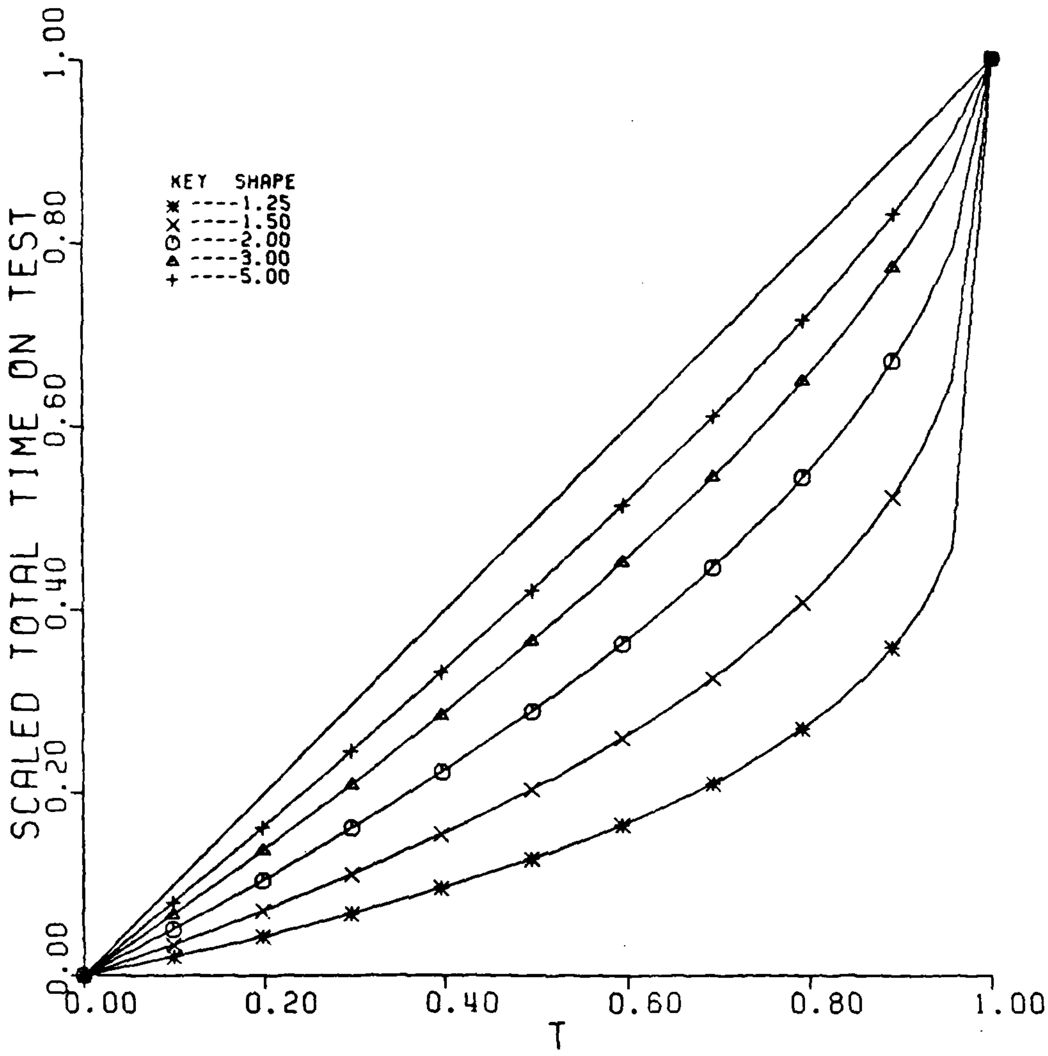


FIGURE 5.1

SCALED TOTAL TIME ON TEST TRANSFORM
FOR GAMMA MODEL.

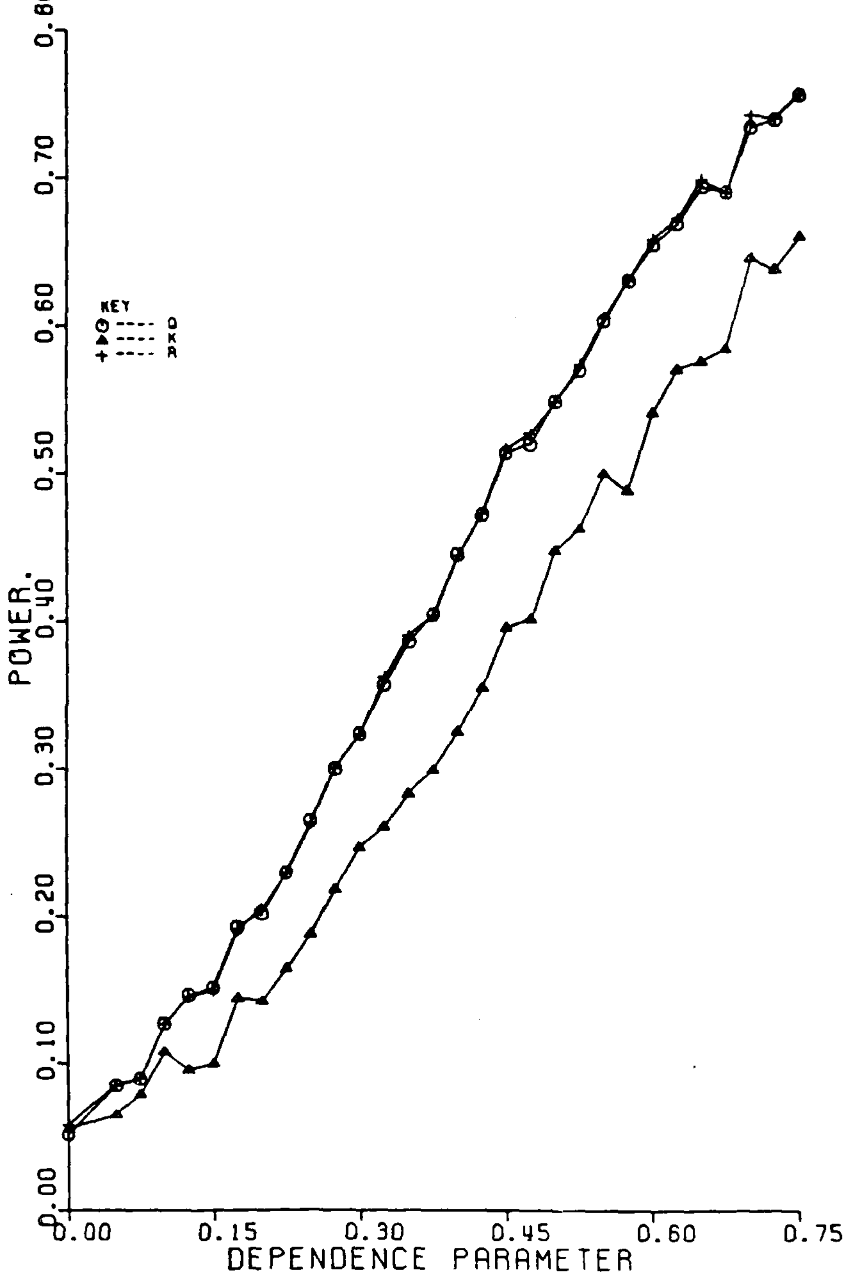


FIGURE 7.1

ESTIMATED POWERS OF TESTS FOR INDEPENDENCE FOR S = 20

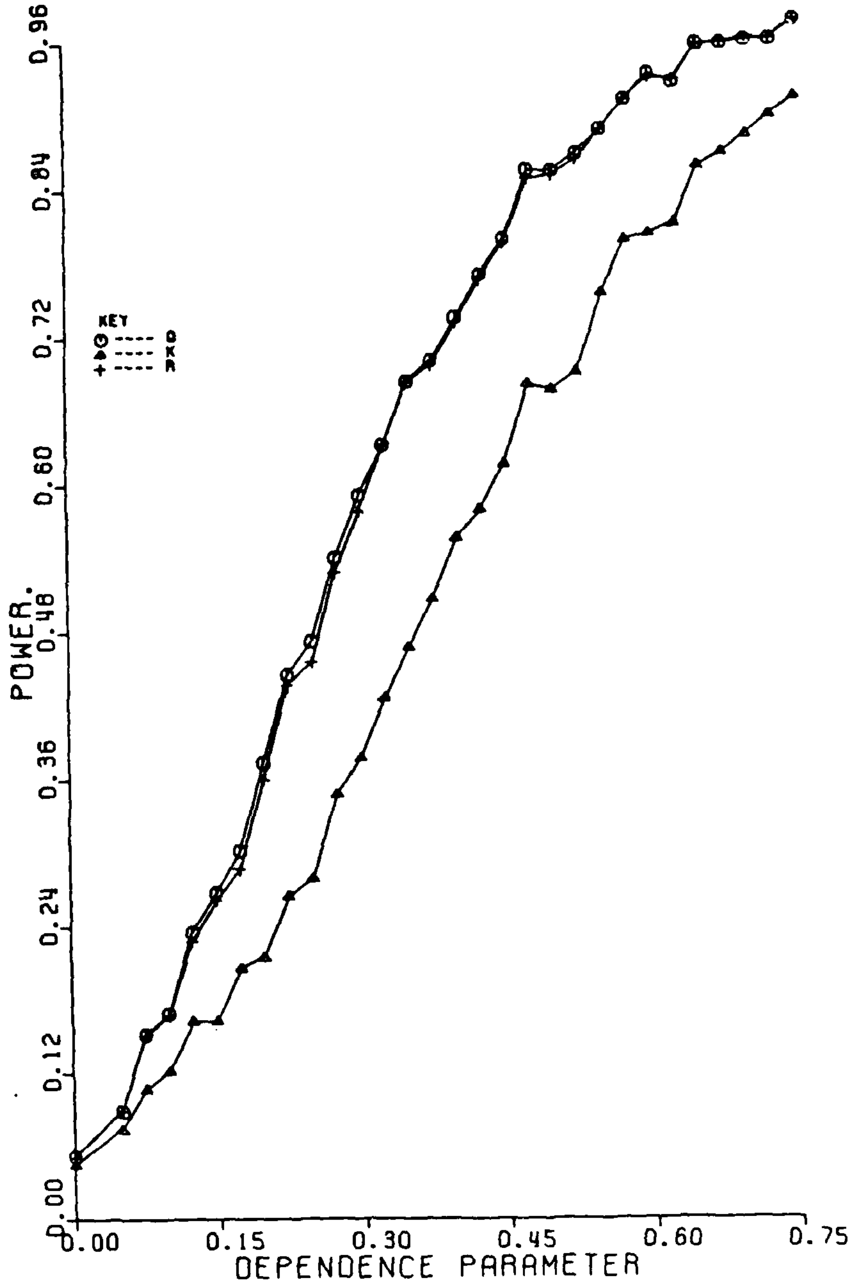


FIGURE 7.2

ESTIMATED POWERS OF TESTS FOR
INDEPENDENCE FOR S = 50

APPENDIX K

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY NA	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE NA	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION The Ohio State University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NM
6c. ADDRESS (City, State, and ZIP Code) 1314 Kinnear Road Columbus, Ohio 43212	7b. ADDRESS (City, State, and ZIP Code) AFOSR/NM Bldg 410 Bolling AFB DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-82-0307
8c. ADDRESS (City, State, and ZIP Code) AFOSR/NM Bldg 410 Bolling AFB DC 20332-6448	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO.: 61102F PROJECT NO.: 2304 TASK NO.: WORK UNIT ACCESSION NO.:

11. TITLE (Include Security Classification) A Random Environmental Stress Model for Competing Risks.

12. PERSONAL AUTHOR(S)
John P. Klein and Sukhoon Lee

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 9-1-82 TO 12-31-87	May 31, 1988	14

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Random environmental effect; dependent competing risks; tpta; total time on test
XXXX	XXXXXXXXXX	XXX	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

A random environmental effects model is proposed for competing risks experiments. The model assumes a random stress, Z, which changes the scale parameter of each of the assumed Weibull times to occurrence of the risks. Some general properties of the model are discussed, and specific properties for a Uniform or Gamma stress model are presented. Estimation of parameters under the Gamma stress model is considered, and a new estimator based on the scaled total time on test transform is presented.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Maj. Brian Woodruff	22b. TELEPHONE (Include Area Code) (202) 767-5027 22c. OFFICE SYMBOL NM

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

A RANDOM ENVIRONMENTAL STRESS MODEL FOR COMPETING RISKS

John P. Klein and Sukhoon Lee

Department of Statistics
The Ohio State University
Columbus, Ohio 43210

ABSTRACT

A random environmental effects model is proposed for competing risks experiments. The model assumes a random stress, Z, which changes the scale parameter of each of the assumed Weibull times to occurrence of the risks. Some general properties of the model are discussed, and specific properties for a Uniform or Gamma stress model are presented. Estimation of parameters under the Gamma stress model is considered, and a new estimator based on the scaled total time on test transform is presented.

INTRODUCTION

The problem of competing risks arises naturally in a number of engineering or biological experiments. In such experiments, for some items put on test, the primary event of interests (such as death, component failure, etc.) is not observable due to the occurrence of some competing risk of removal from the study (such as censoring, failure from a different component, etc.). Competing risks arise in an engineering context in analyzing data from

(a) series systems,

(b) field tests of equipment with a fixed test time and a random or staggered entry into the study, or

(c) systems with multiple failure modes.

Competing risks arise in biological applications in analyzing data from

A RANDOM ENVIRONMENTAL STRESS MODEL FOR COMPETING RISKS

John P. Klein and Sukhoon Lee

Department of Statistics
The Ohio State University
Columbus, Ohio 43210

TR 326

August, 1985

(a) clinical trials with a fixed trial duration and staggered entry

(b) clinical trials with some patients withdrawing from the trial prior to response

(c) studies of the time to death from a variety of causes

A common assumption made in analyzing competing risks experiments is that the potential (unobservable) times to occurrence of the competing risks are independent. This assumption is not testable due to the identifiability problem. That is, for any dependent competing risks model, there exists an independent competing risks model which yields the same observables. (See Basu and Klein (1982) for details.) However, Moeschberger and Klein (1984) show that an investigator can be appreciably misled in modeling competing risks by erroneously assuming indepencence.

In this paper we present a model for dependence between the various risks by assuming that dependence is due to some common environmental factor which effects the potential times to occurrences of each risk. In section 2 we present the model and study its properties for bivariate series and parallel systems. In section 3, we consider estimation of the model parameters for competing risks systems.

2. THE MODEL

For simplicity we shall consider the problem of bivariate systems and discuss our model in terms of engineering applications. We assume that under ideal, controlled conditions, as one may encounter in the laboratory in the testing or design stage of development, the time to failure of the two components, to be linked in a system, are X_0 and Y_0. We suppose that under these conditions, X_0, Y_0 have survival functions F_0, G_0 on $[0, \infty)$. We assume that both X_0 and Y_0 follow a Weibull form with parameters (η_1, λ_1) and (η_2, λ_2), respectively, That is, $F_0(x) = \exp(-\lambda_1 x^{\eta_1})$. The Weibull distribution, which may have increasing ($\eta > 1$), decreasing ($\eta < 1$) or constant failure rate ($\eta = 1$) has been shown experimentally to provide a reasonable fit to many different types of survival data. (See Bain (1978)). We now link the two components into a system in such a way that under ideal lab conditions the two components are independent.

Now suppose that the above system (X_0, Y_0) is put into operation under usage conditions. We suppose that under such conditions the effect of the environment is to degrade or improve each component by the same random amount. That is, the effect of the environment is to select a random factor, Z, from some distribution, H, which changes the maginal survival functions of the two components to F_0^Z and G_0^Z. A value of Z less than one means that component reliabilities are simultaneously improved, while a value of Z greater than one implies a joint degradation. The resulting joint reliability of the two components' lifetimes, (X,Y) in the operating environment is

$$F(x,y) = E[\exp(-Z(\lambda_1 x^{\eta_1} + \lambda_2 y^{\eta_2}))]. \tag{2.1}$$

This model has been proposed by Lindley and Singpurwalla (1984) in the reliability context when F_0, G_0 are exponential and H( ) follows a gamma distribution. This basic dependence structure was also proposed by Clayton (1978) to model associations in bivariate survival data, and later by Oakes (1982) to model bivariate survival data. Hutchinson (1982) proposed a similar model when H( ) has a gamma distribution and $F_0(t) = G_0(t) = \exp(-t^{\eta})$.

The model described above for a general distribution of the environmental stress has a particular dependence structure which we summarize in the following lemmas.

Lemma 1. Let (X,Y) follow the model (2.1) where Z is a positive random variable with finite $(\frac{r}{\eta_1} + \frac{s}{\eta_2})^{th}$ inverse moment. Then

$$E(X^r Y^s) = \lambda_1^{-r/\eta_1} \lambda_2^{-s/\eta_2} \Gamma(1 + r/\eta_1)\, \Gamma(1 + s/\eta_2)\, E(Z^{-(r/\eta_1 + s/\eta_2)}) \tag{2.2}$$

The proof follows by noting that, given Z = z, (X,Y) are independent Weibulls with parameters $(\eta_1, \lambda_1 z)$ and $(\eta_2, \lambda_2 z)$, respectively and $E(X^r|Z=z) = \lambda_1^{-r/\eta_1} z^{-r/\eta_1} \Gamma(1+ r/\eta_1)$ with a similar expression for Y^s. When the appropriate moments exist, we have

(A) $E(X) = E(X_0)\, E(Z^{-1/\eta_1})$,

(B) $V(X) = E(X_0^2)\, Var(Z^{-1/\eta_1}) + E(Z^{-1/\eta_1})^2\, Var(X_0)$,

(C) $Cov(X,Y) = E(X_0)\, E(Y_0)\, Cov(Z^{-1/\eta_1}, Z^{-1/\eta_2})$ which is greater than 0.

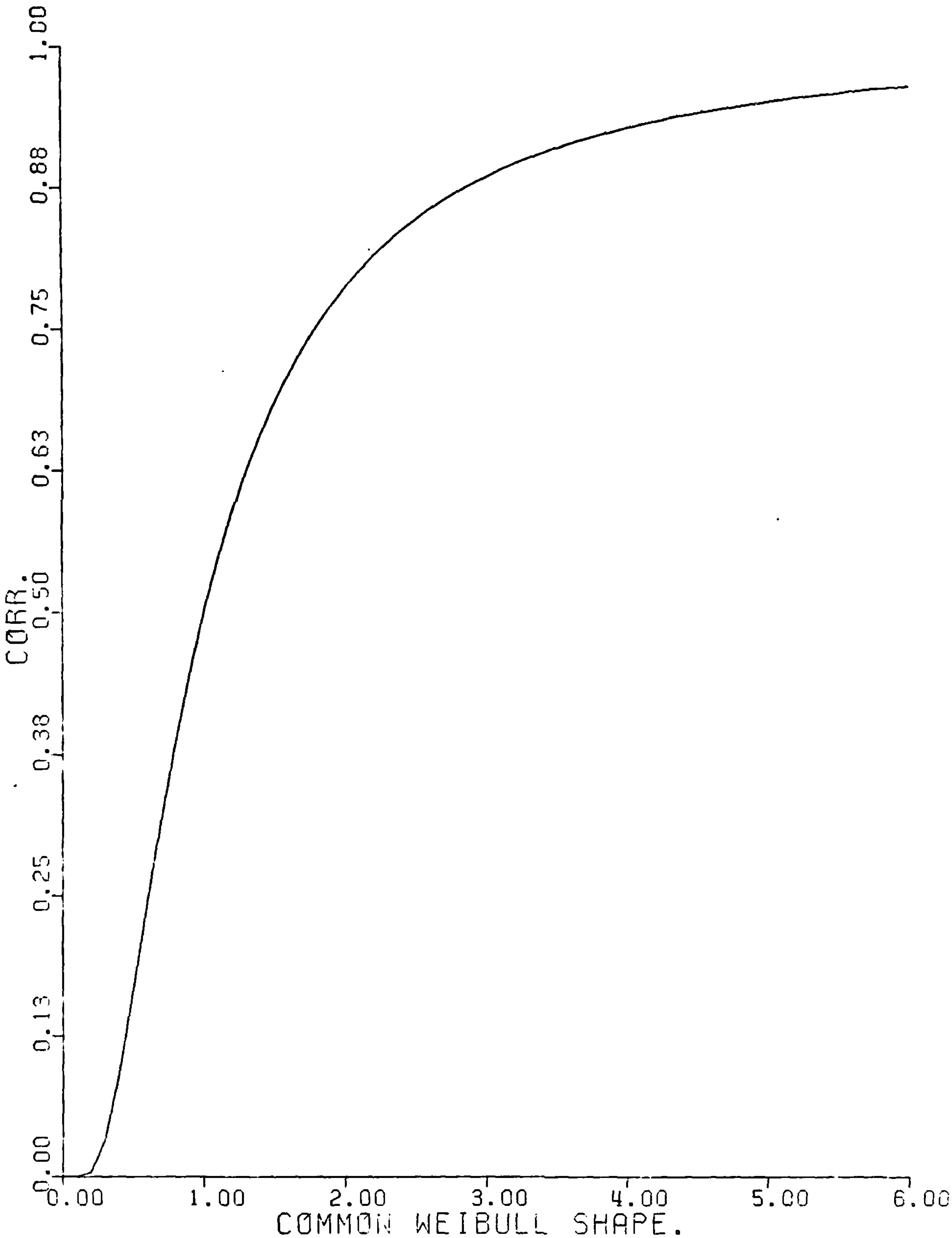
If $\eta_1 = \eta_2 = \eta$ then the correlation between (X,Y) is

$$\rho = \frac{\Gamma(1+ 1/\eta)^2\, Var(Z^{-1/\eta})}{Var(Z^{-1/\eta})\, \Gamma(1+2/\eta) + (\Gamma(1+2/\eta) - \Gamma(1+1/\eta)^2)\, E(Z^{-1\eta})^2}.$$

In this case the correlation is bounded above by $\Gamma(1+1/\eta)^2 / \Gamma(1+2/\eta)$. Figure 1 shows the maximal correlation as a functcion of η for $\eta \in (0, 10)$. Note that this maximal correlation is an

FIGURE 1

UPPER BOUND ON MAXIMAL CORRELATION FOR RANDOM ENVIORNMENT MODEL.



increasing function of η. One can also show that F(x, y) is positive quadrant dependent for any η_1, η_2.

Exact expressions for competing risks quantities of interest can be computed when a particular model is assumed for the distribution of Z. We shall consider the gamma and uniform models. Consider first the gamma model with $h_Z(z) = b^a z^{a-1} \exp(-bz)/\Gamma(a)$, $z > 0$. For this model, the joint survival function is

$$F(x,y) = \frac{b^a}{[b+\lambda_1 x^{\eta_1} + \lambda_2 y^{\eta_2}]^a} \tag{2.3}$$

which is a bivariate Burr Distribution (see Takahasi (1965)), the marginal distributions are univariate Burr distributions with

$$E(X) = (\lambda_1/b)^{-1/\eta_1}\Gamma(1+1/\eta_1)\Gamma(a-1/\eta_1)\,\Gamma(a), \text{ if } a > 1/\eta,$$

$$Var(X) = (\lambda_1/b)^{-2/\eta_1}\left\{\frac{\Gamma(1+2/\eta_1)\Gamma(a-2/\eta_1)}{\Gamma(a)} - \left[\frac{\Gamma(1+1/\eta_1)\Gamma(a-1/\eta_1)}{\Gamma(a)}\right]^2\right\}, \text{ if } a > 2/\eta_1$$

with similar expressions for E(Y), Var(Y). The covariance of (X,Y) is

$$Cov(X,Y) = (\lambda_1/b)^{-1/\eta_1}(\lambda_2/b)^{-1/\eta_2}\Gamma(1+1/\eta_1)\,\Gamma(1+1/\eta_2)\left\{\frac{\Gamma(a-1/\eta_1-1/\eta_2)}{\Gamma(a)} - \frac{\Gamma(a-1/\eta_2)\,\Gamma(a-1/\eta_2)}{\Gamma(a)}\right\}$$

for $a > 1/\eta_1 + 1/\eta_2$. For the gamma model, the reliability function for a bivariate series system is given by

$$R_S(t) = (1+(\lambda_1/b)t^{\eta_1} + (\lambda_2/b)t^{\eta_2})^{-a}, \tag{2.4}$$

and for a parallel system by

$$R_p(t) = (1+(\lambda_1/b)t^{\eta_1} + (1+(\lambda_2/b)t^{\eta_2})^{-a} - (1+(\lambda_1/b)t^{\eta_2} + (\lambda_2/b)t^{\eta_2})^{-a} \tag{2.5}$$

Figures 2A-E are plots of the series system reliability for $\lambda_1 = 1$, $\lambda_2 = 2$ and several combinations of η_1, η_2. Each figure shows the reliablity for a = 1/2, 1, 2, 4, and the independent Weibull model. In all cases, b = 1. For these figures we note that for fixed λ_1, λ_2 η_1, η_2, t, the series system reliability is a decreasing function of the shape parameter a. Figures 3A-E are plots of the parallel system reliability (2.5) for the above parameters. Again, the reliability is a decreasing function of a. Also in both the series and parallel system reliability, the shape of the reliability function is quite different from that encountered under independence.

The gamma model is a reasonable model for the environmental stress due to its flexibility and the tractability of the model in obtaining close form solutions for the relevant quantities and in estimating parameters. However, in some cases, such as when the operating environment is always more severe than the laboratory environment, the support of H may be restricted to some fixed interval. A possible model for such an environmental stress is the uniform distribution over [a,b]. For this model, the joint survival function is

$$F(x,y) = \frac{[\exp(-b(\lambda_1 x^{\eta_1} + \lambda_2 y^{\eta_2})) - \exp(-a(\lambda_1 x^{\eta_1} + \lambda_2 y^{\eta_2}))]}{(b-a)(\lambda_1 x^{\eta_1} + \lambda_2 y^{\eta_2})} \tag{2.6}$$

$$E(X) = \lambda_1^{-1/\eta_1} \Gamma(1+1/\eta_1)\, \eta_1 \left(b^{(\eta_1-1)/\eta_1} - a^{(\eta_1-1)/\eta_1}\right) / \{(\eta_1-1)(b-a)] \qquad \text{if } \eta_1 \neq 1$$

$$= \ln(b/a)/[\lambda_1(b-a)] \qquad \text{if } \eta_1 = 1,$$

$$Var(X) = \frac{\eta_1 \lambda_1^{-2/\eta_1}}{(b-a)} \{ \Gamma(1+2/\eta_1)\, \eta_1 \left(b^{(\eta_1-2)/\eta_1} - a^{(\eta_1-2)/\eta_1}\right)$$

$$- \frac{\Gamma(1+1/\eta_1)^2\, \eta_1 \left(b^{(\eta_1-1/\eta_1)} - a^{(\eta_1-1/\eta_1)}\right)^2}{(\eta_1-1)^2(b-a)} \} \qquad \text{if } \eta_1 \neq 1,2$$

$$2/(\lambda_1{}^2 ab) - \ln(b/a)^2/[(b-a)\,\lambda_1]^2 \qquad \text{if } \eta_1 = 1$$

$$\lambda_1^{-1}\left[\frac{\ln(b/a)}{(b-a)} - \frac{\Pi}{(b^{1/2} + a^{1/2})^2}\right] \qquad \text{if } \eta_1 = 2$$

FIGURE 2 A

SERIES SYSTEM RELIABILITY UNDER GAMMA(A,1) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0,\quad \lambda_2=2.0,\quad \eta_1=1.0,\quad \eta_2=1.0.$

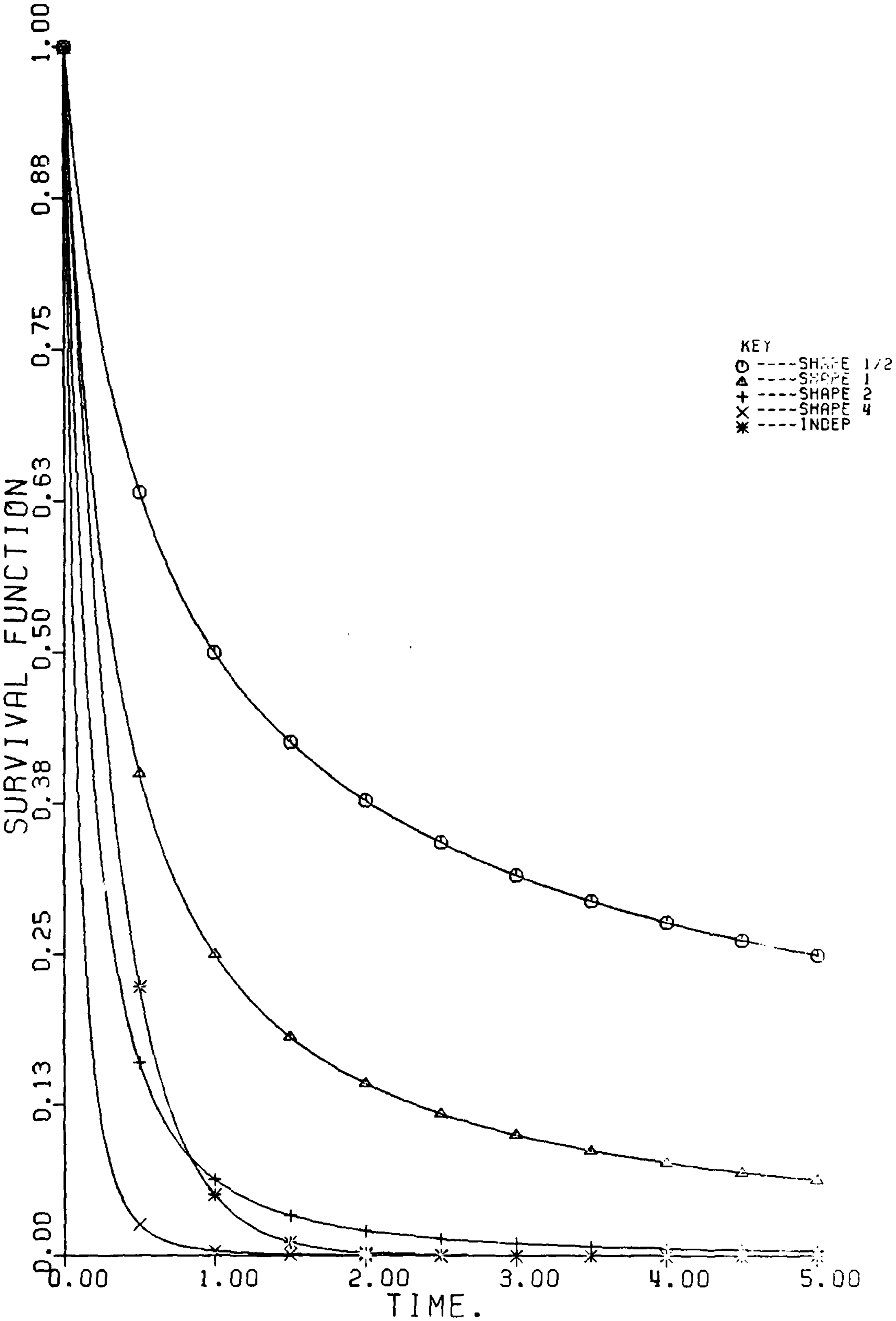


FIGURE 2 B

SERIES SYSTEM RELIABILITY UNDER GAMMA(A,1) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=1.0, \quad \eta_2=2.0.$

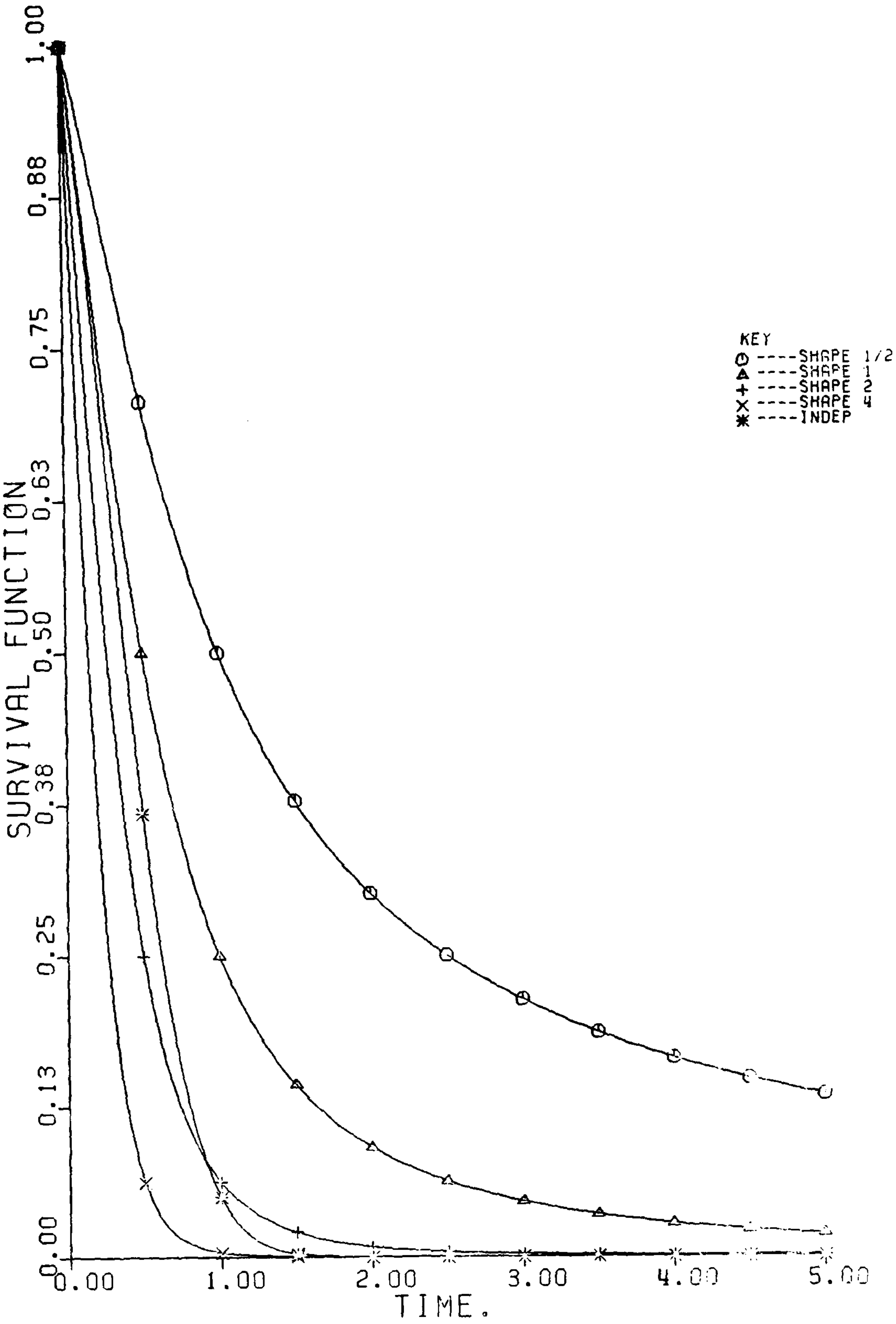


FIGURE 2 C

SERIES SYSTEM RELIABILITY UNDER GAMMA(A,1) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=2.0, \quad \eta_2=2.0.$

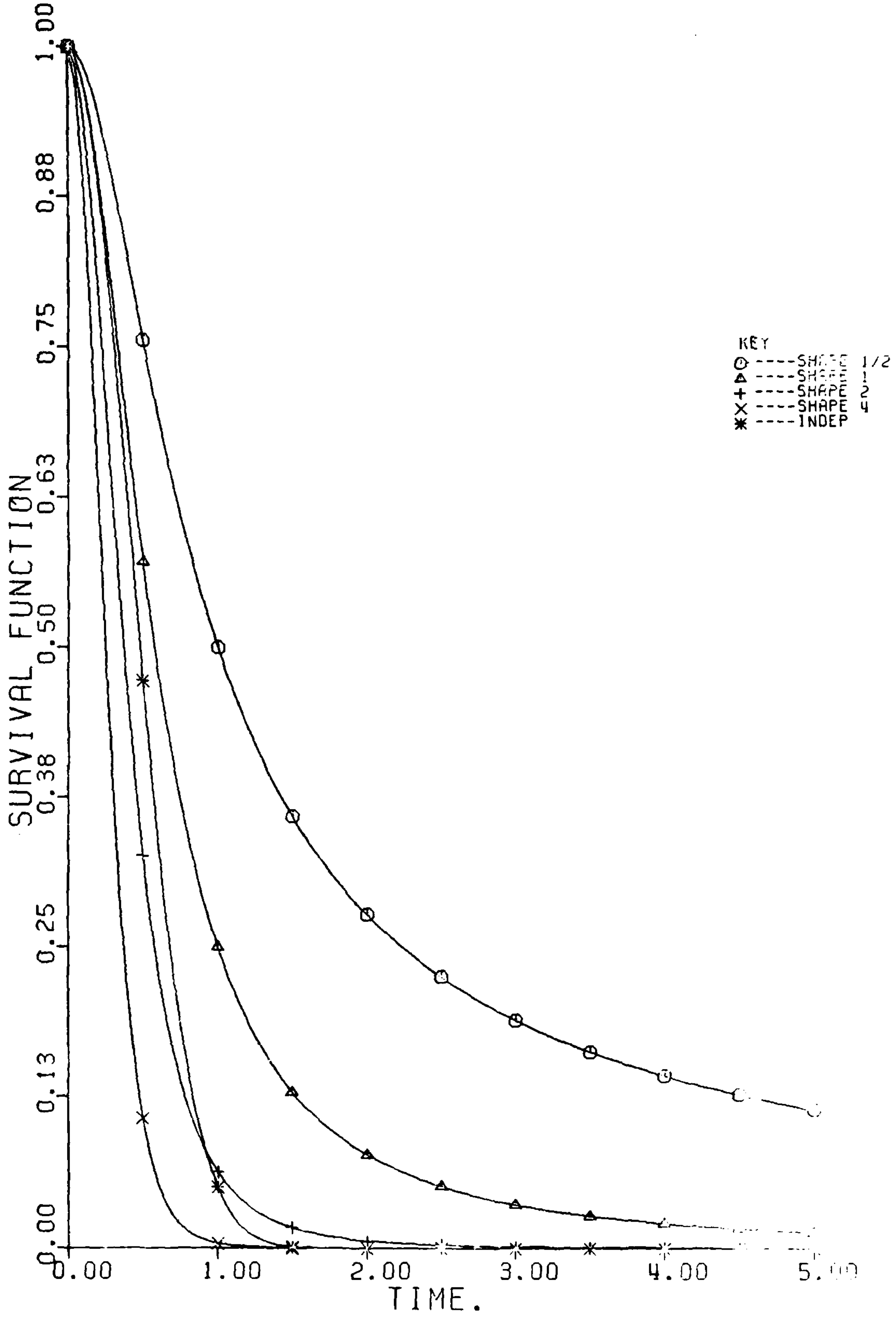


FIGURE 2 D

SERIES SYSTEM RELIABILITY UNDER GAMMA(A,1) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=1/2, \quad \eta_2=1/2.$

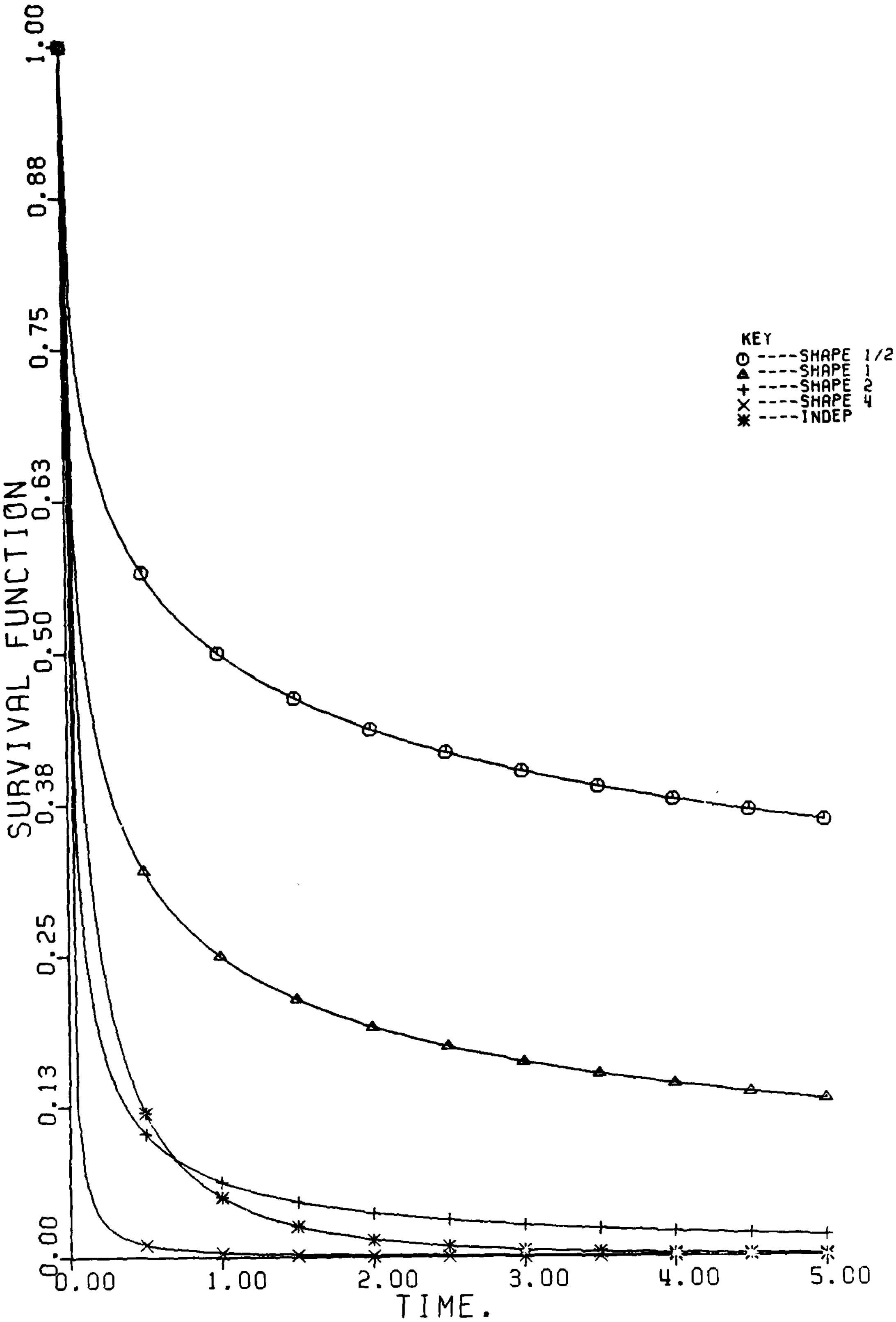


FIGURE 2 E

SERIES SYSTEM RELIABILITY UNDER GAMMA(A,1) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=1/2, \quad \eta_2=2.0.$

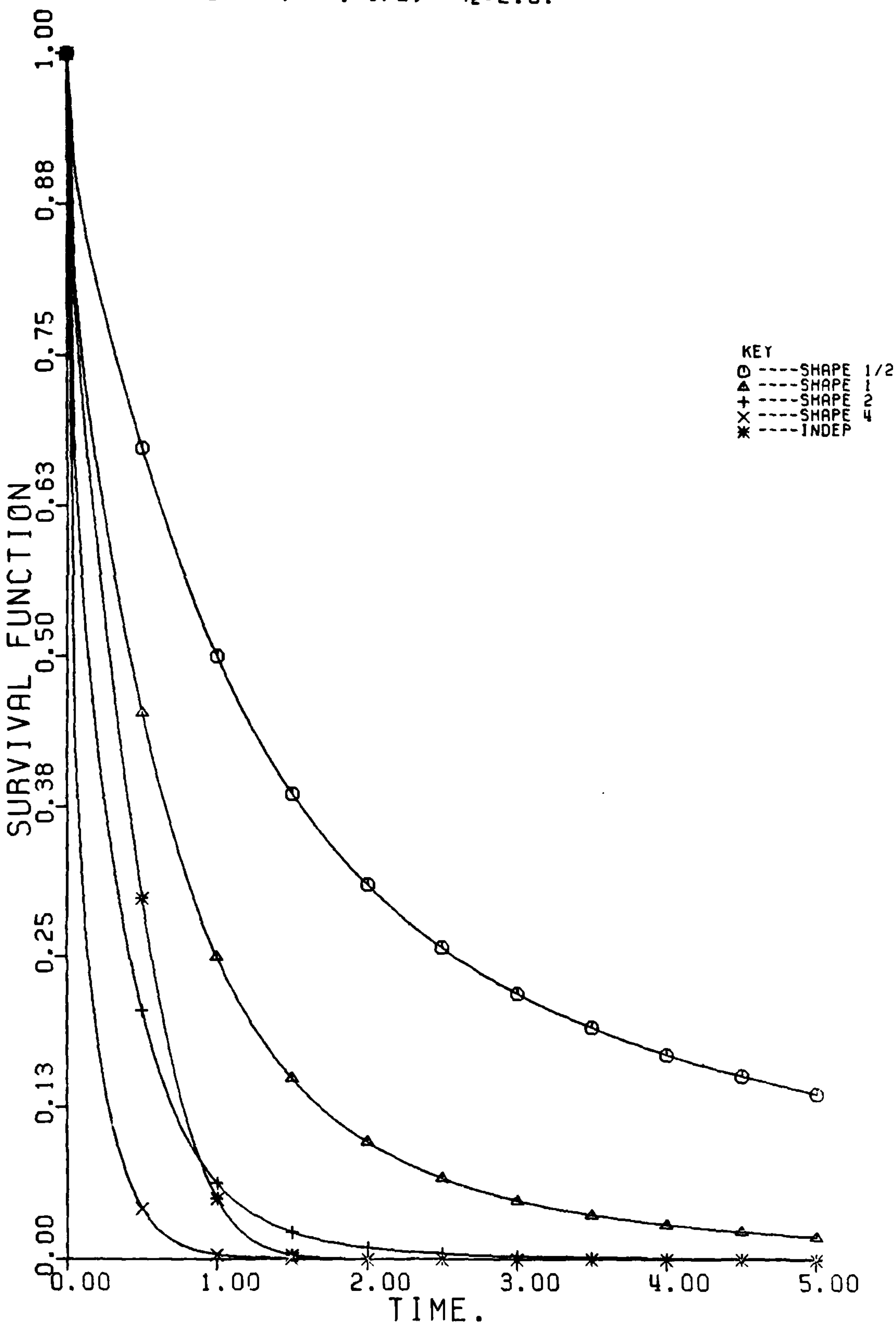


FIGURE 3 A

PARALLEL SYSTEM RELIABILITY UNDER GAMMA(A,1) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0$, $\lambda_2=2.0$, $\eta_1=1.0$, $\eta_2=1.0$.

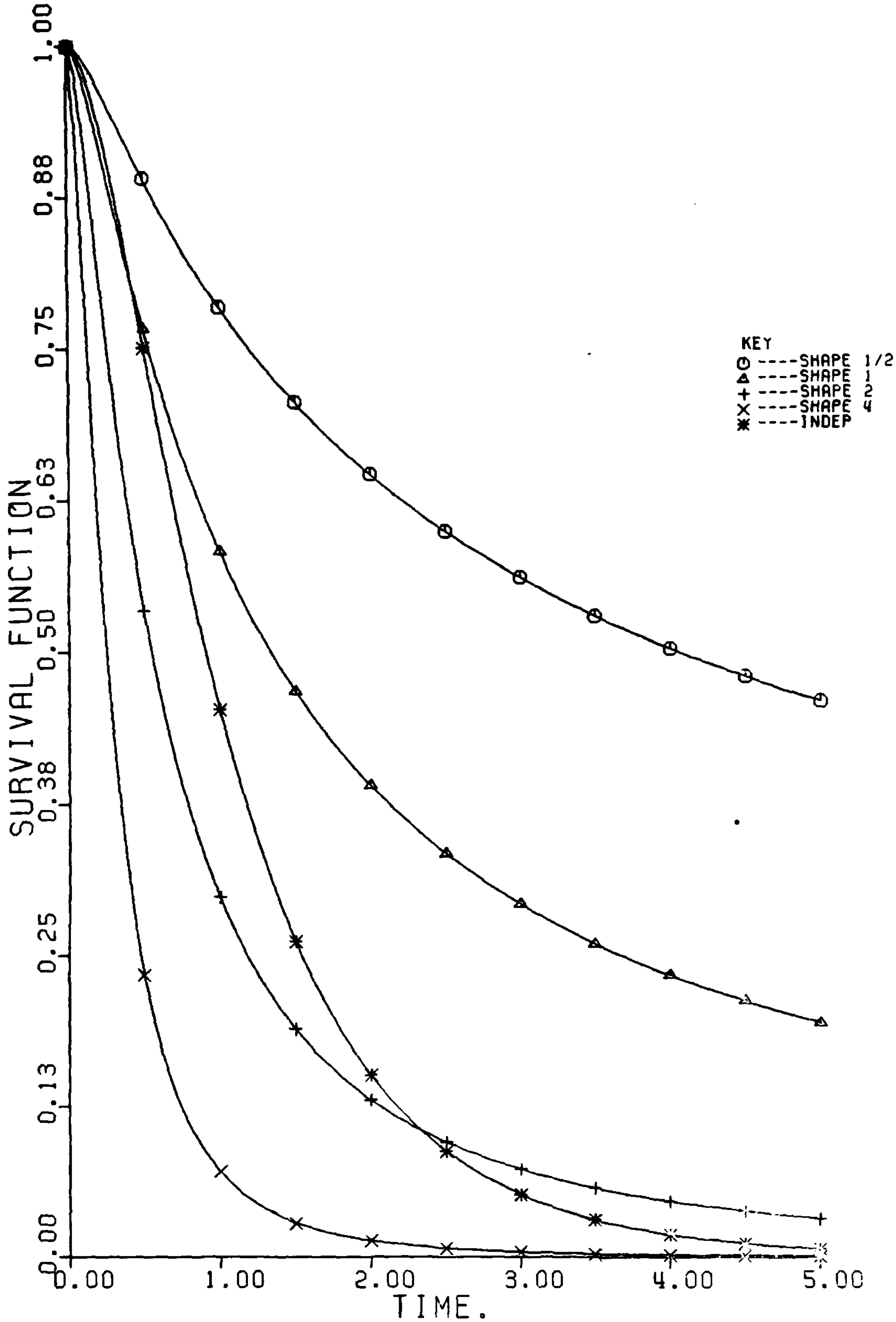


FIGURE 3 B

PARALLEL SYSTEM RELIABILITY UNDER GAMMA(A,1) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=1.0, \quad \eta_2=2.0.$

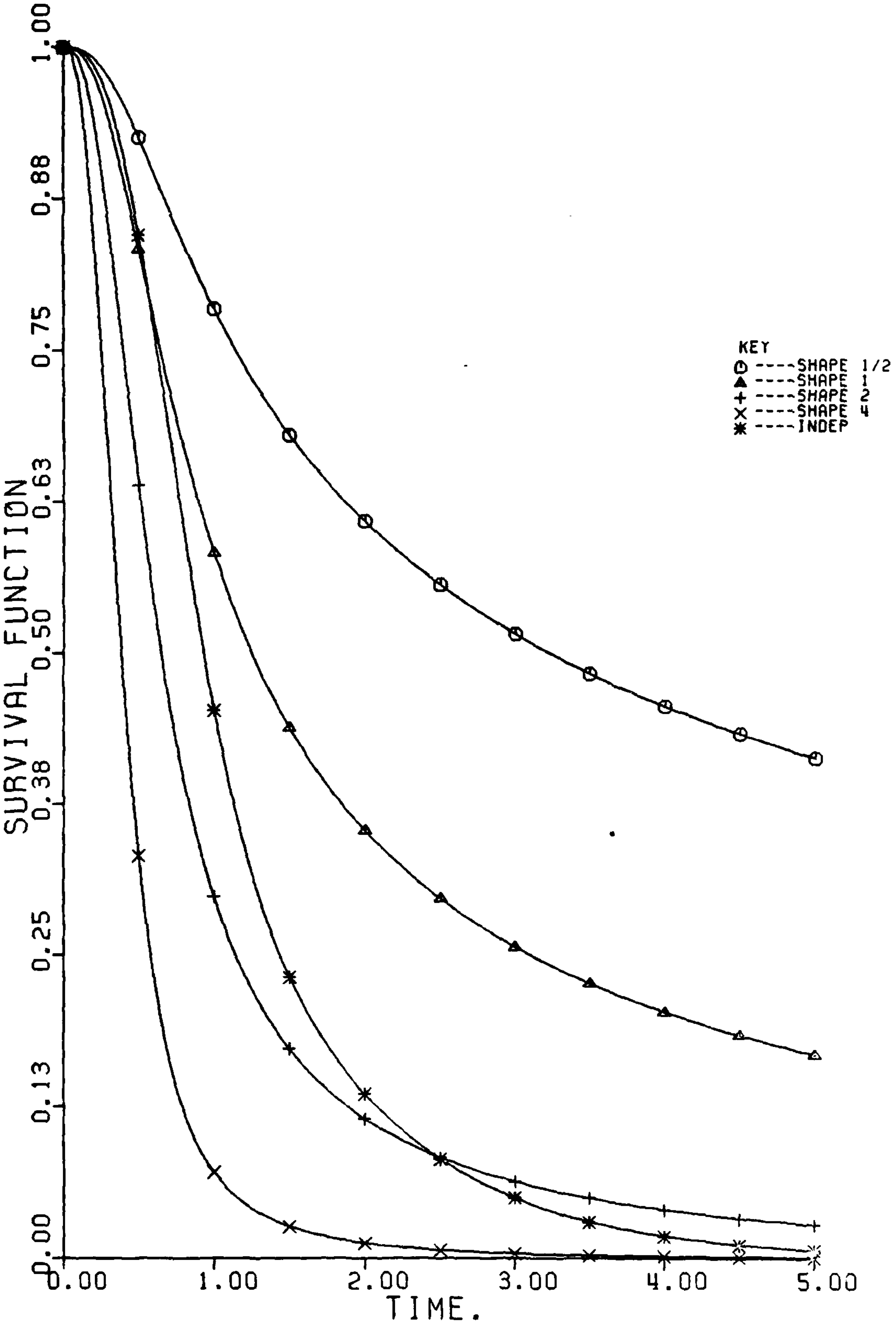


FIGURE 3 C

PARALLEL SYSTEM RELIABILITY UNDER GAMMA(A,1) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=2.0, \quad \eta_2=2.0.$

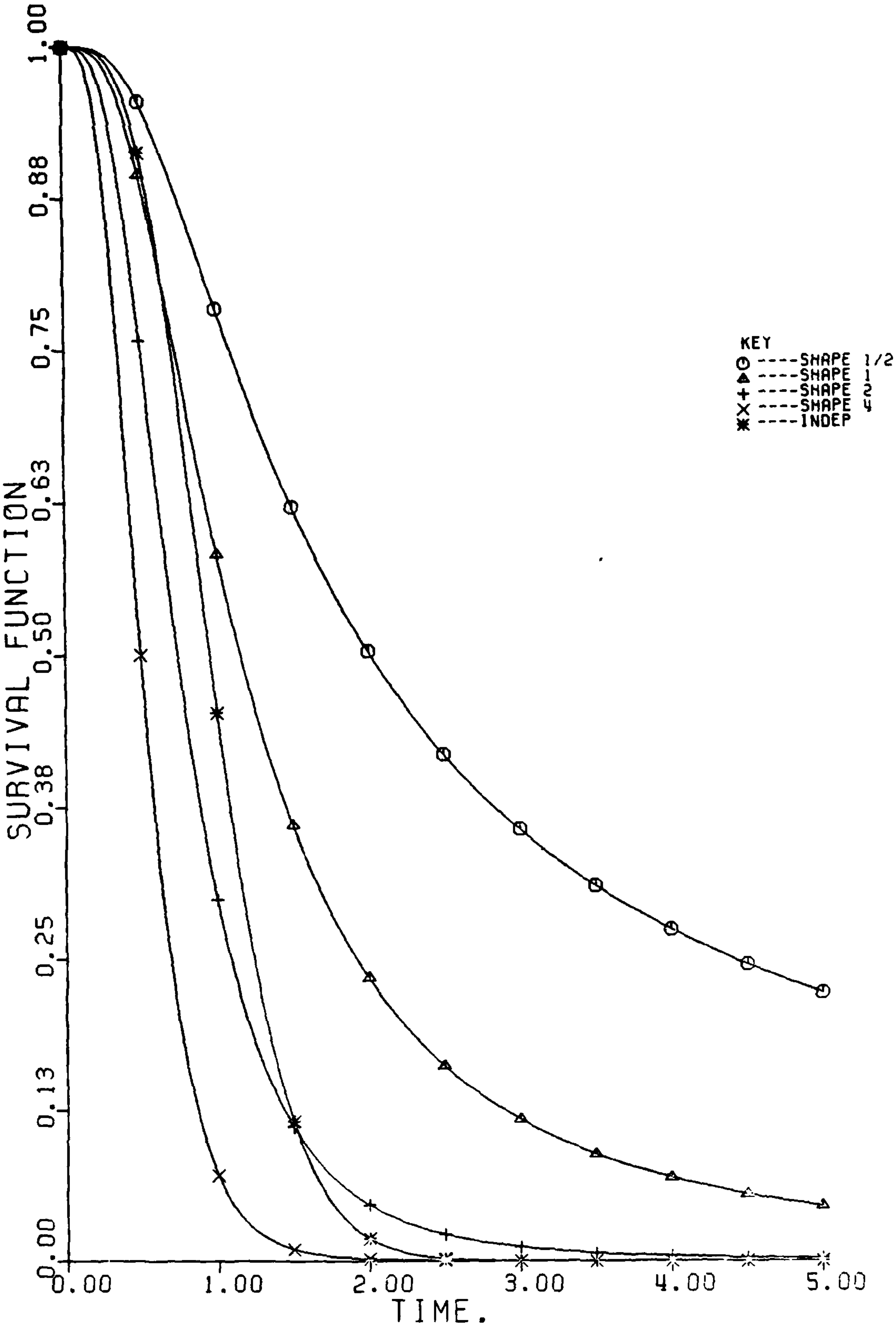


FIGURE 3 D

PARALLEL SYSTEM RELIABILITY UNDER GAMMA(A,1) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0$, $\lambda_2=2.0$, $\eta_1=1/2$, $\eta_2=1/2$.

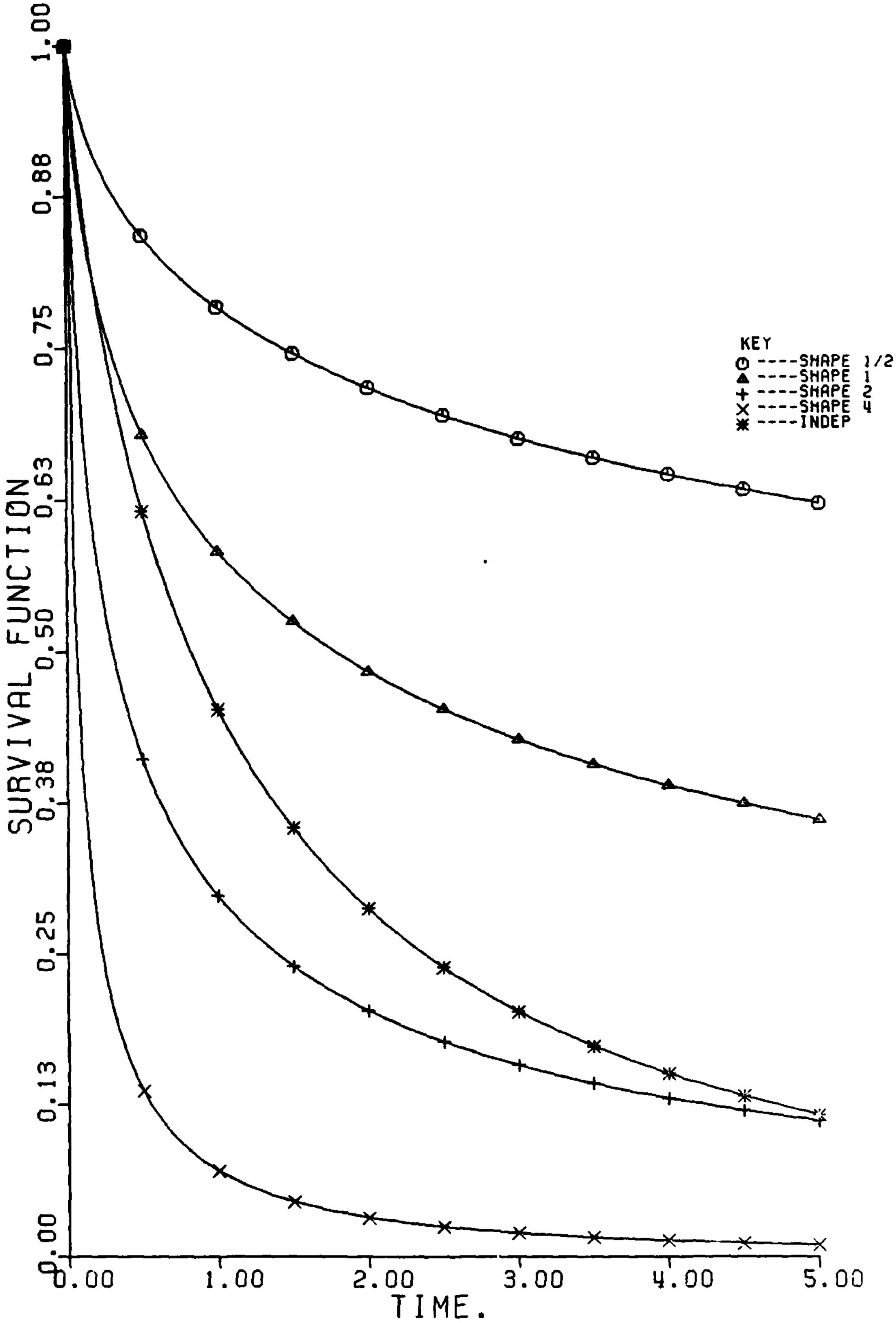
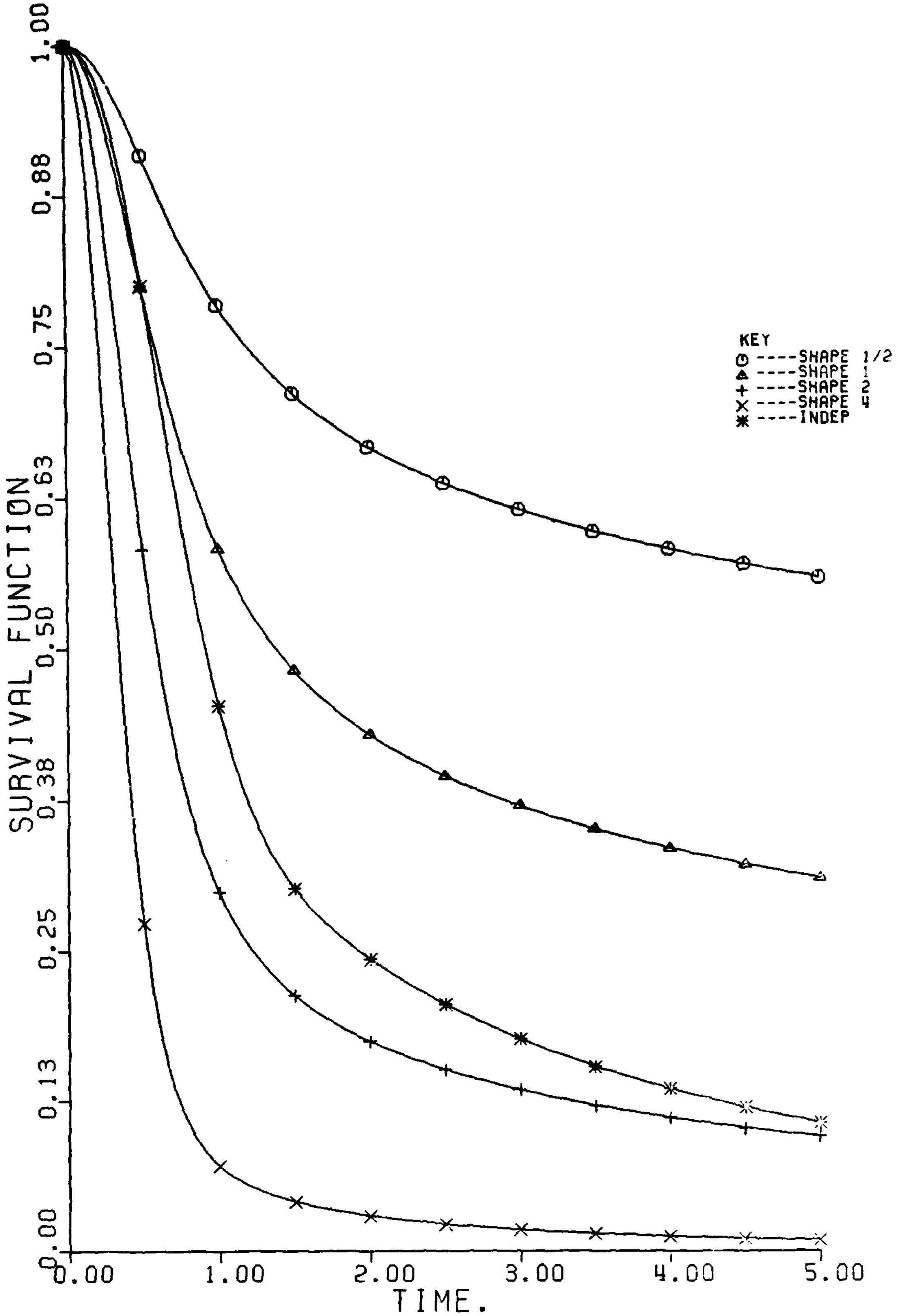


FIGURE 3 E

PARALLEL SYSTEM RELIABILITY UNDER GAMMA(A,1) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0$, $\lambda_2=2.0$, $\eta_1=1/2$, $\eta_2=2.0$.



and

$$\mathrm{Cov}(X,Y) = \frac{\Gamma(1+1/\eta_1)\,\Gamma(1+1/\eta_2)}{\lambda_1^{1/\eta_1}\lambda_2^{1/\eta_1}}\Bigg\{\frac{\eta_1\eta_2}{(\eta_1\eta_2-\eta_1-\eta_2)}\,\frac{\left(b^{\frac{\eta_1\eta_2-\eta_1-\eta_2}{\eta_1\eta_2}} - a^{\frac{\eta_1\eta_2-\eta_1-\eta_2}{\eta_1\eta_2}}\right)}{(b-a)} -$$

$$\frac{\eta_1\eta_2}{(\eta_1-1)(\eta_2-1)}\,\frac{\left(b^{\frac{\eta_1-1}{\eta_1}} - a^{\frac{\eta_1-1}{\eta_1}}\right)\left(b^{\frac{\eta_2-1}{\eta_2}} - a^{\frac{\eta_2-1}{\eta_2}}\right)}{(b-a)^2}\Bigg\} \text{if } \eta_1 \neq 1,\ \eta_2 \neq 1,\ 1/\eta_1 + 1/\eta_2 \neq 1,$$

$$\frac{\Gamma(1+1/\eta_1)\,\Gamma((2\eta_1-1)/\eta_1)}{\lambda_1^{1/\eta_1}\,\lambda_2^{\eta_1/(\eta_1-1)}}\Bigg[\frac{\ln(b/a)}{(b-a)} - \frac{\eta_1^2}{(\eta_1-1)}\,\frac{\left(b^{\frac{\eta_1-1}{\eta_1}} - a^{\frac{\eta_1-1}{\eta_1}}\right)}{(b-a)^2}\left(b^{1/\eta_1} - a^{1/\eta_1}\right)\Bigg]$$

$$\text{if } 1/\eta_1 + 1/\eta_1 = 1$$

$$\frac{\Gamma(1+1/\eta_i)}{\lambda_i^{1/\eta_i}\lambda_{i'}}\Bigg[\frac{\eta_i\eta_{i'}}{(\eta_i\eta_{i'}-\eta_i-\eta_{i'})(b-a)}\left(b^{\frac{\eta_i\eta_{i'}-\eta_i-\eta_{i'}}{\eta_i\eta_{i'}}} - a^{\frac{\eta_i\eta_{i'}-\eta_i-\eta_{i'}}{\eta_i\eta_{i'}}}\right) - \frac{\eta_i}{(\eta_i-1)}\,\frac{\left(b^{\frac{\eta_i-1}{\eta_i}} - a^{\frac{\eta_i-1}{\eta_i}}\right)\ln(b/a)}{(b-a)^2}$$

$$\text{if } \eta_i \neq 1,\ \eta_{i'} = 1$$

$$\frac{1}{(\lambda_1\ \lambda_2)}\left[\frac{1}{(ab)} - \frac{\ln(b/a)^2}{(b-a)^2}\right] \text{ if } \eta_1 = \eta_2 = 1$$

For this model, the reliability function for a series system is

$$R_s(t) = \underline{[\exp(-b(\lambda_1 t^{\eta_1} + \lambda_2 t^{\eta_2}) - \exp[-a(\lambda_1 t^{\eta_1} + \lambda_2 t^{\eta_2})]]} \qquad (2.7)$$

$$(b-a)(\lambda_1 t^{\eta_1} + \lambda_2 t^{\eta_2})$$

and for a parallel system is

$$R_p(t) = \frac{[\exp(-b(\lambda_1 t^{\eta_1}) - \exp(-a\lambda_1 t^{\eta_1})}{(b-a)\lambda_1 t^{\eta_1}} + \frac{[\exp(-b\lambda_2 t^{\eta_2}) - \exp(-a\lambda_2 t^{\eta_2})}{(b-a)\lambda_2 t^{\eta_2}} - Rs(t)$$

Figures 4A-E show the reliability for a series system and figures 5A-E for a parallel system under the uniform model for various combinations of $\lambda_1, \lambda_2, \eta_1, \eta_2$ a,b. Notice that when A = .25, B = .75, which corresponds to an operating environment which is less severe than the test environment, the system reliability is greater than that expected under independence, while when (a,b) = (1.25, 1.75) or (1., 2), which corresponds to an environment more severe than the test environment, the system reliability is smaller. Also when the (a,b) contains 1, which corresponds to an environment which incurs the possibility of no differential effect from that found in the laboratory, there is little difference in the dependent and independent system reliability.

3. Estimation of Parameters Under Gamma Model

Consider the model (2.3) with $\eta_1 = \eta_2 = \eta$. For this model, the reliablity for a series system is

$$R_s(t) = \left(1 + \frac{(\lambda_1+\lambda_2)}{b} t^{\eta}\right)^{-a}. \qquad (3.1)$$

Notice that this model depends only on two parameters $\theta = (\lambda_1 + \lambda_2)/b$ and a so that if we had data only from systems on test in the operating environment, the only identifiable parameters are a, θ,

FIGURE 4 A

SERIES SYSTEM RELIABILITY UNDER UNIF (A,B) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=1.0, \quad \eta_2=1.0.$

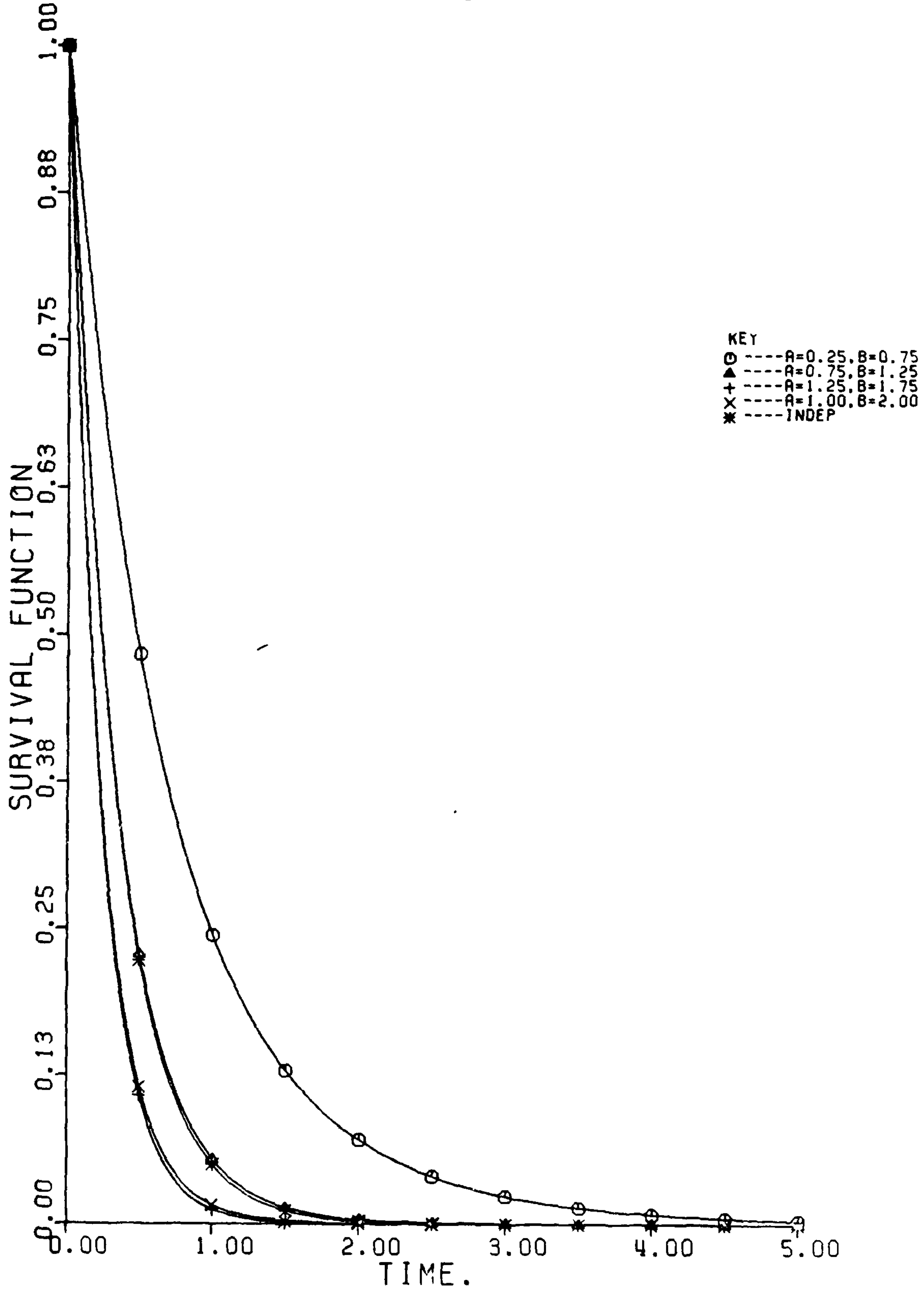


FIGURE 4 B

SERIES SYSTEM RELIABILITY UNDER UNIF (A,B) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=1.0, \quad \eta_2=2.0.$

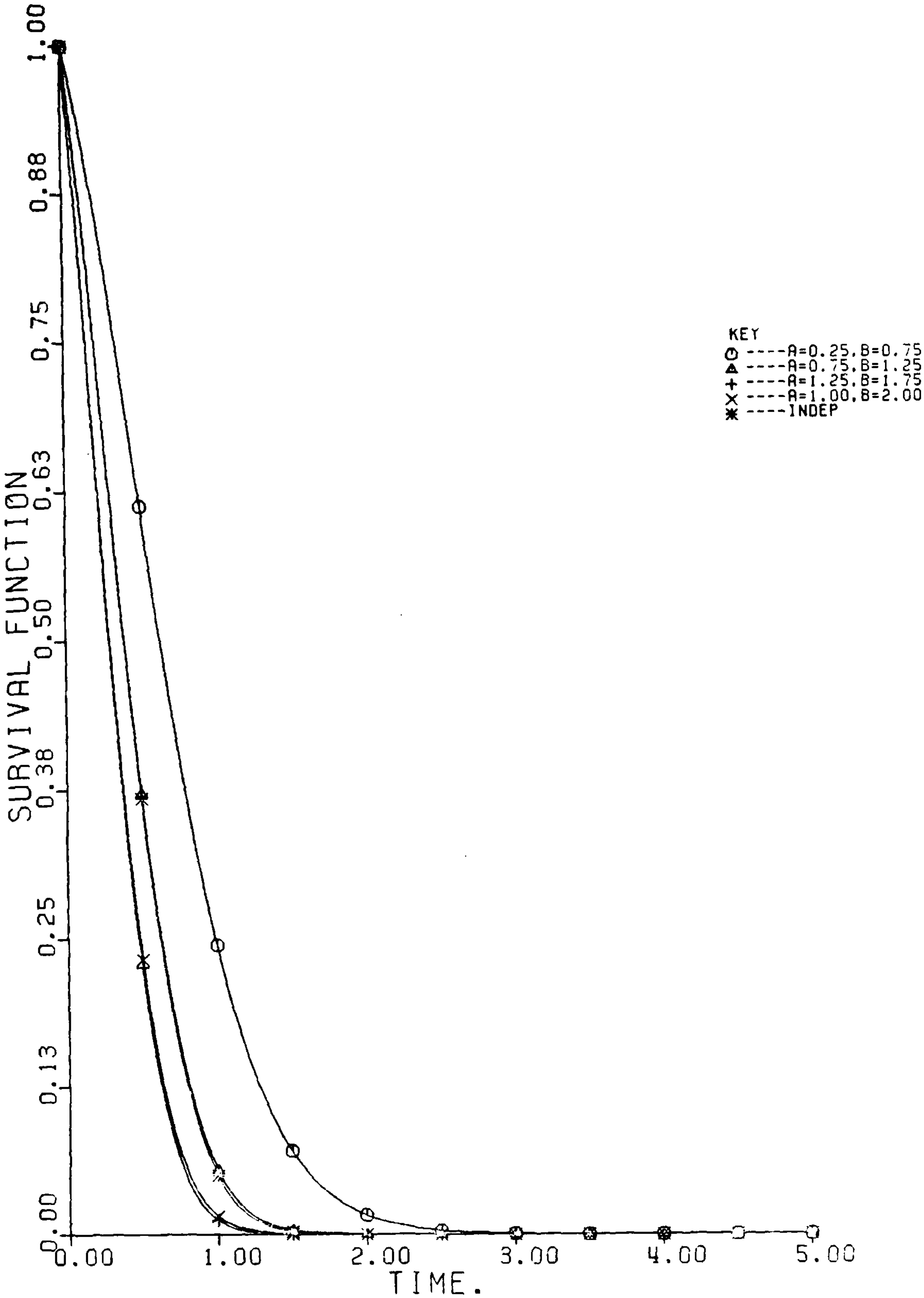


FIGURE 4 C

SERIES SYSTEM RELIABILITY UNDER UNIF (A,B) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=2.0, \quad \eta_2=2.0.$

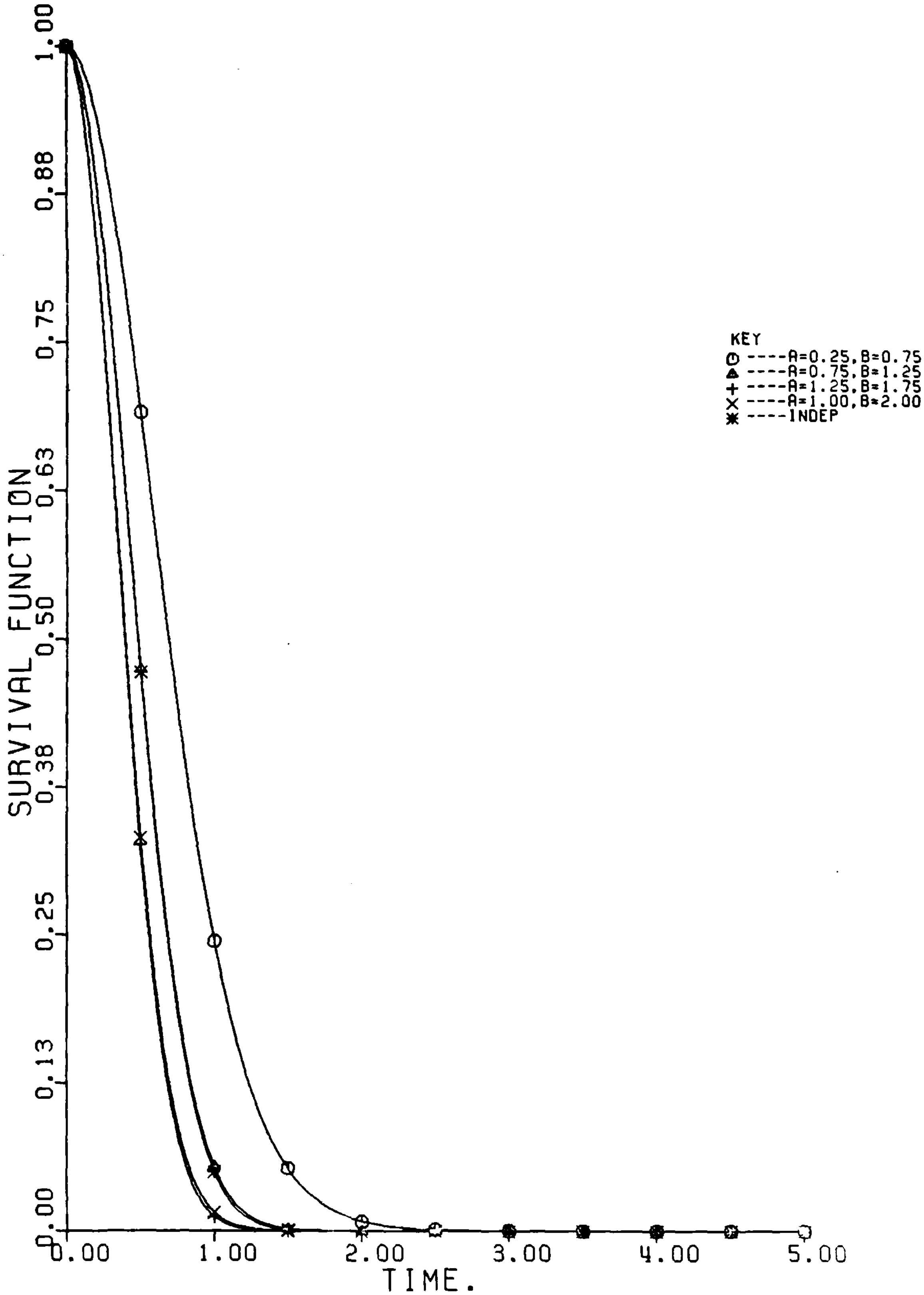


FIGURE 4 D

SERIES SYSTEM RELIABILITY UNDER UNIF (A,B) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=1/2, \quad \eta_2=1/2.$

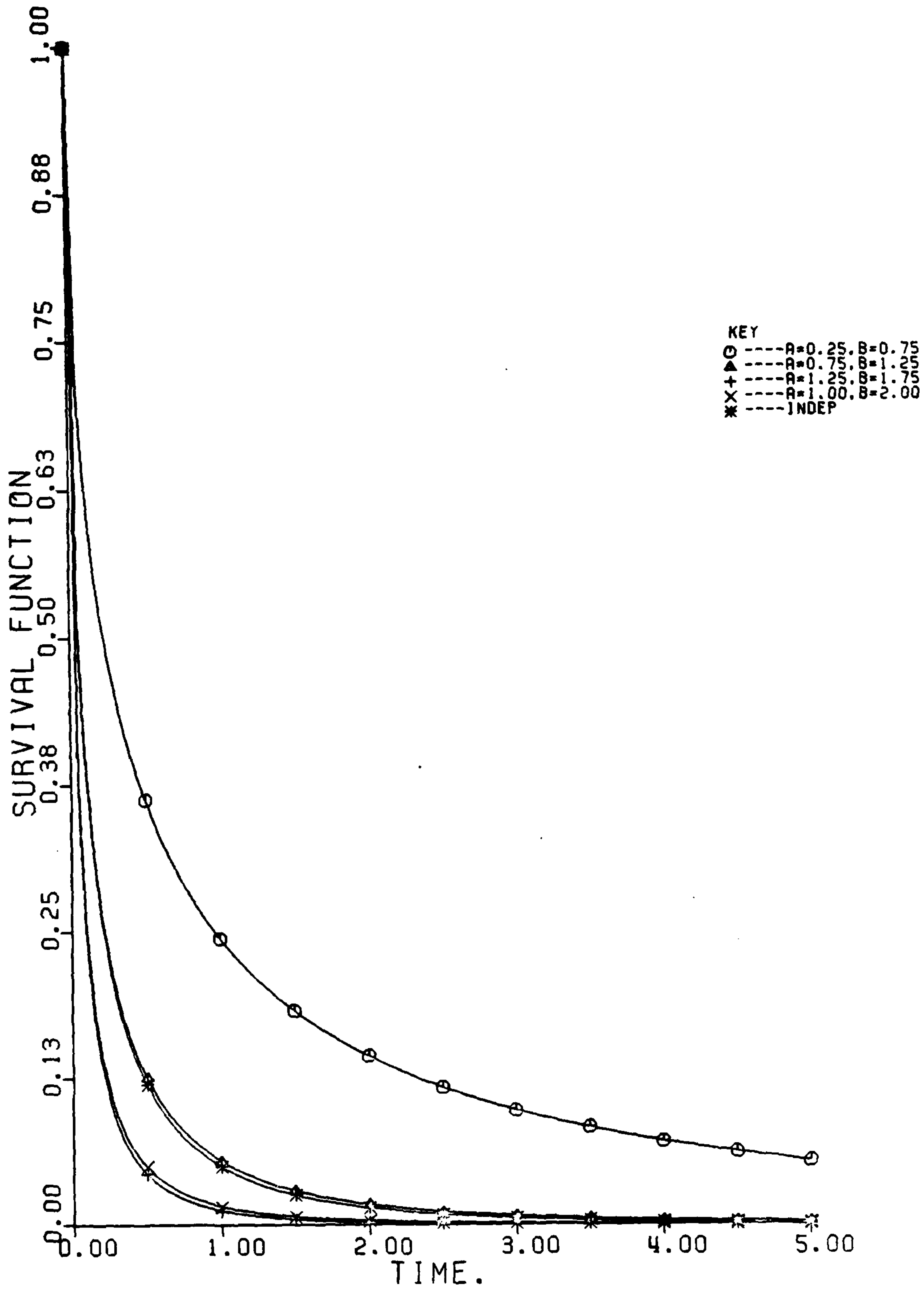


FIGURE 4 E

SERIES SYSTEM RELIABILITY UNDER UNIF (A,B) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=1/2, \quad \eta_2=2.0.$

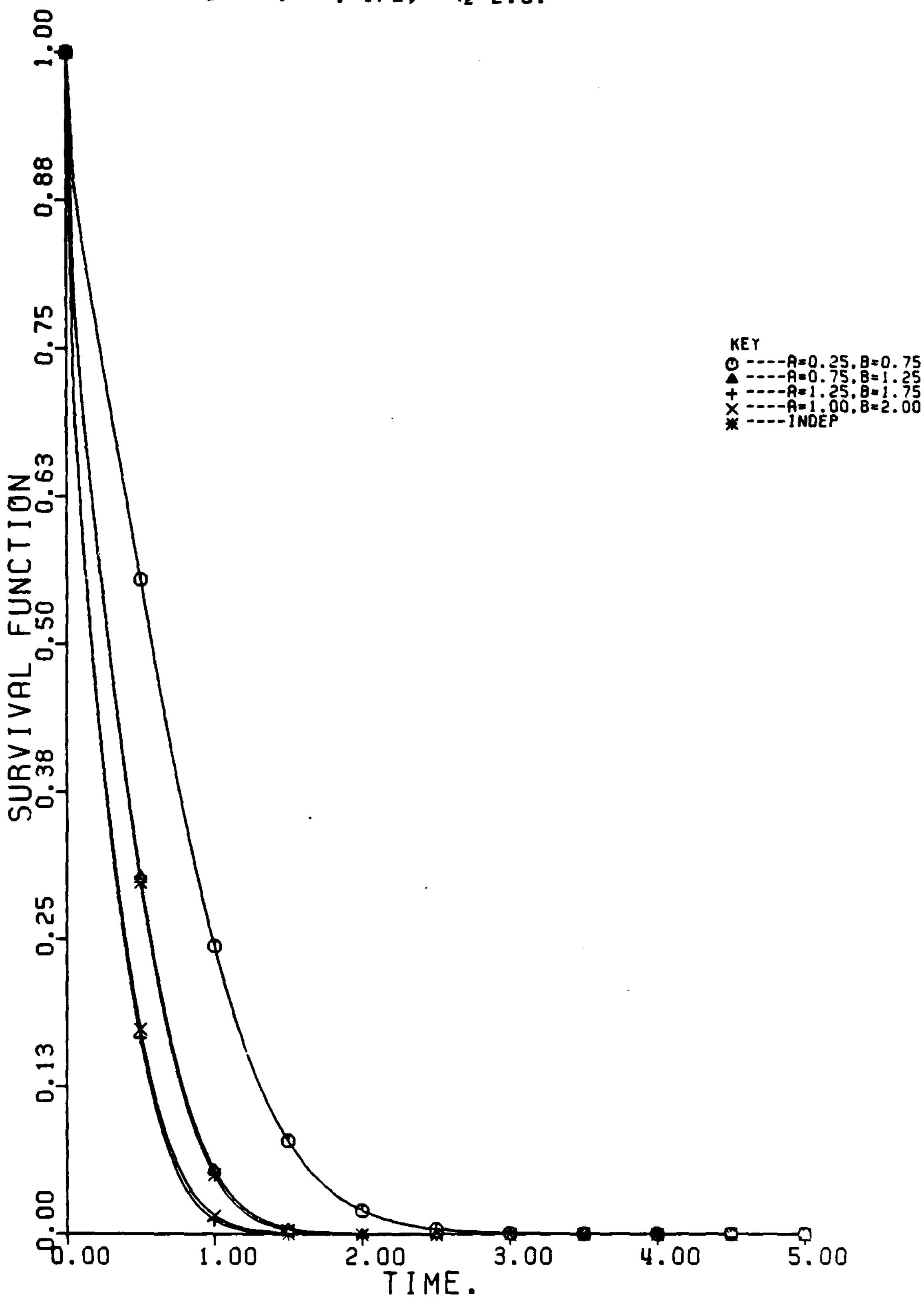


FIGURE 5 A

PARALLEL SYSTEM RELIABILITY UNDER UNIF (A,B) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0$, $\lambda_2=2.0$, $\eta_1=1.0$, $\eta_2=1.0$.

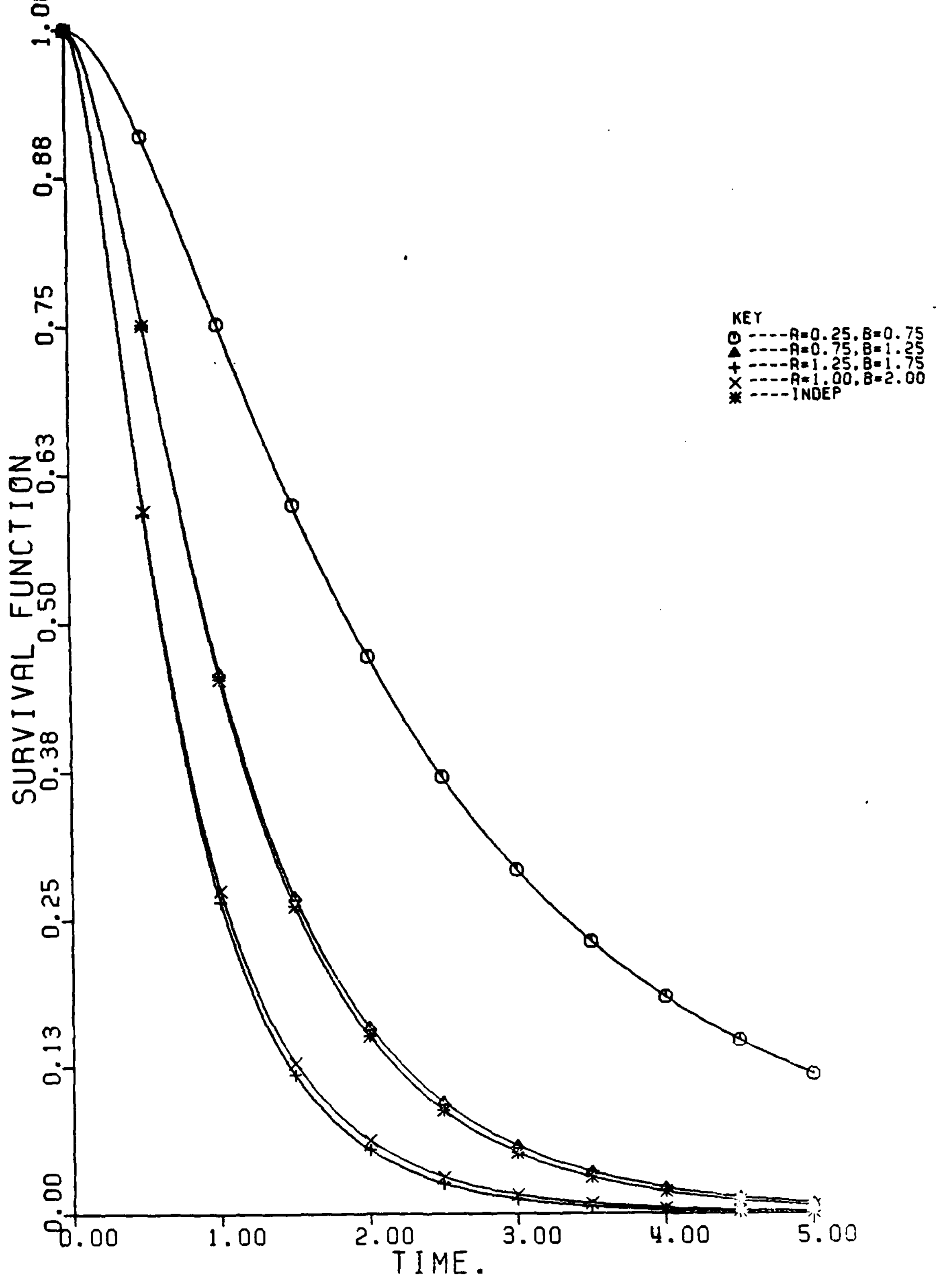


FIGURE 5 B

PARALLEL SYSTEM RELIABILITY UNDER UNIF (A,B) MODEL
FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0$, $\lambda_2=2.0$, $\eta_1=1.0$, $\eta_2=2.0$.

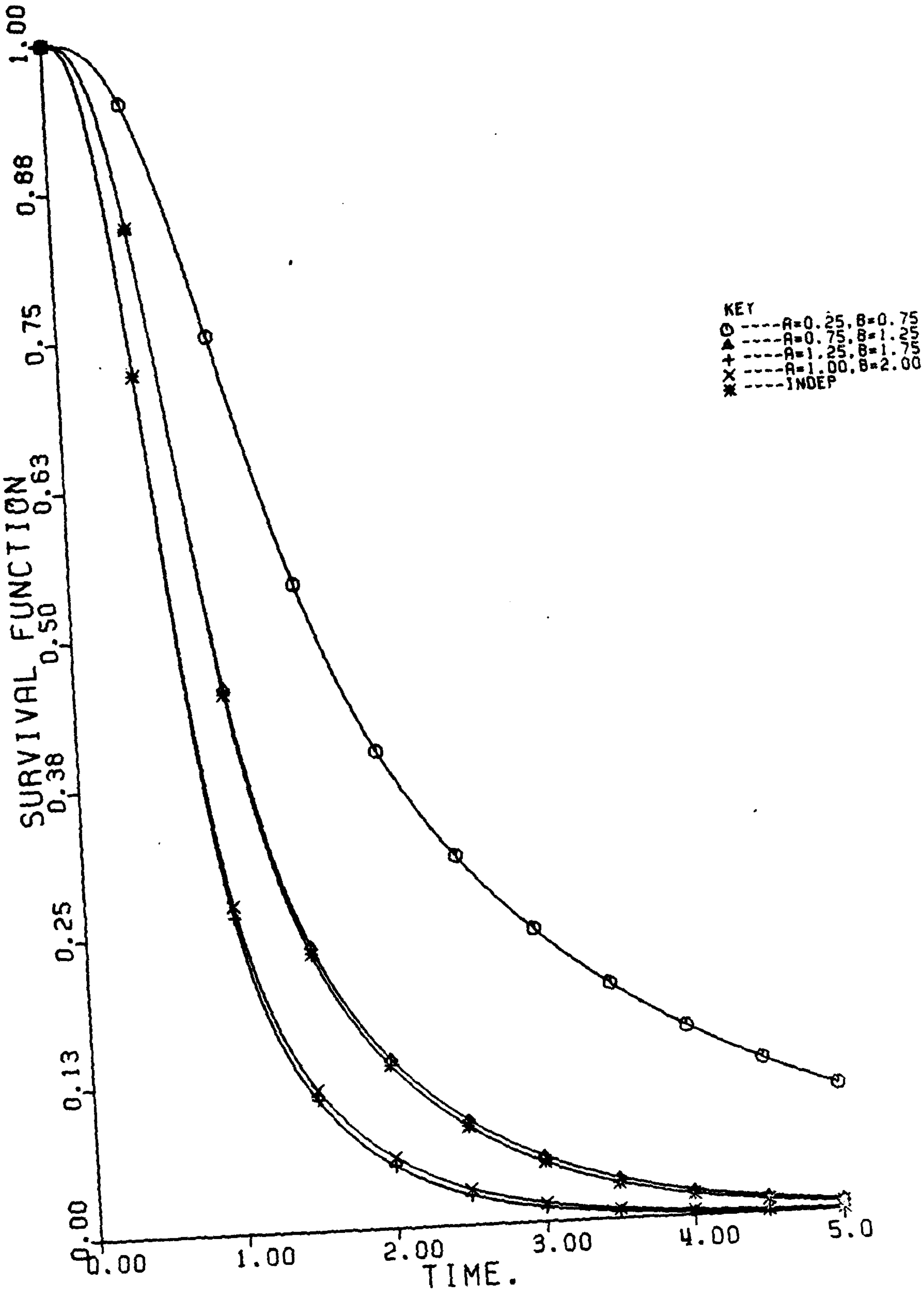


FIGURE 5 C

PARALLEL SYSTEM RELIABILITY UNDER UNIF (A,B) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=2.0, \quad \eta_2=2.0.$

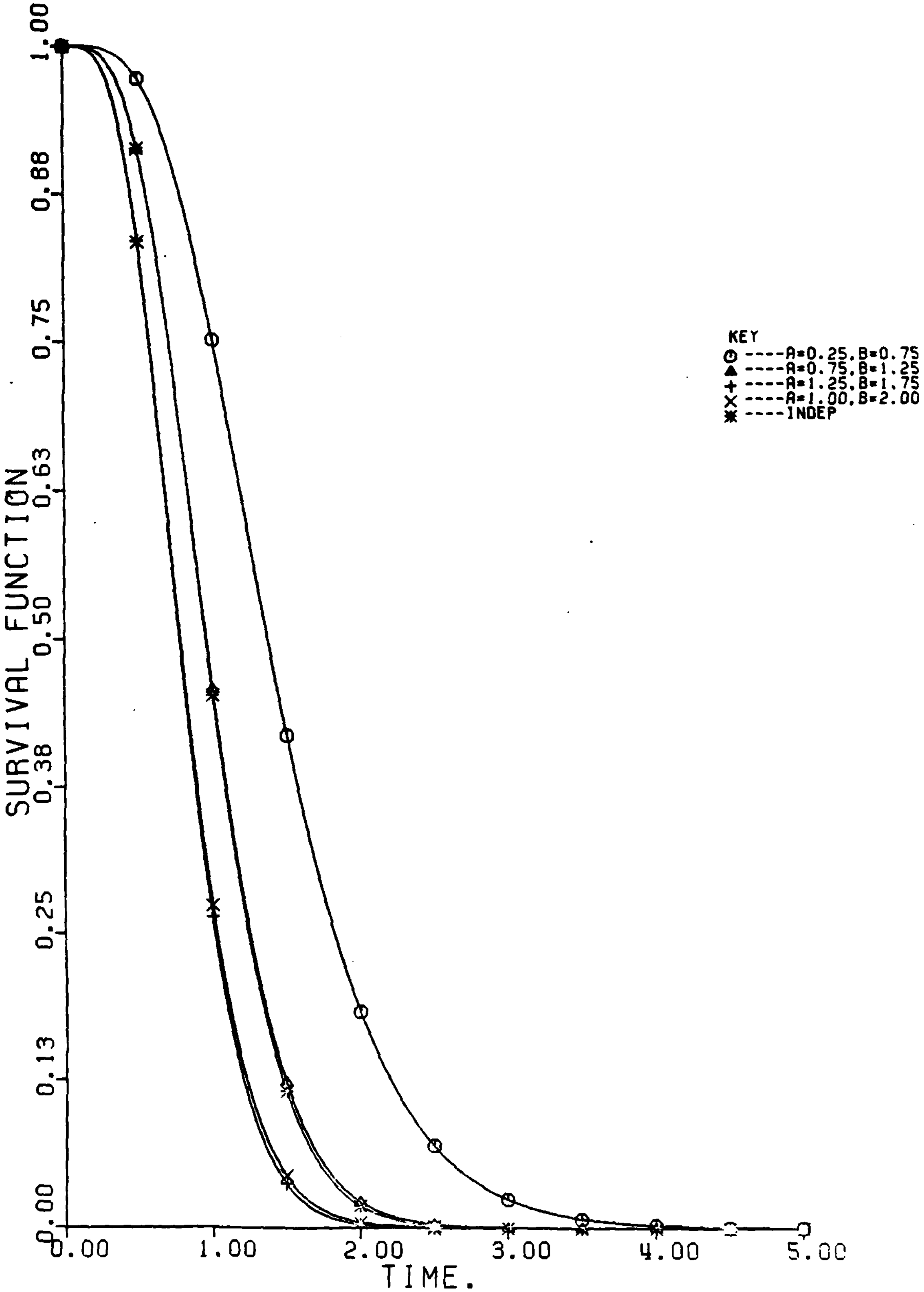


FIGURE 5 D

PARALLEL SYSTEM RELIABILITY UNDER UNIF (A,B) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=1/2, \quad \eta_2=1/2.$

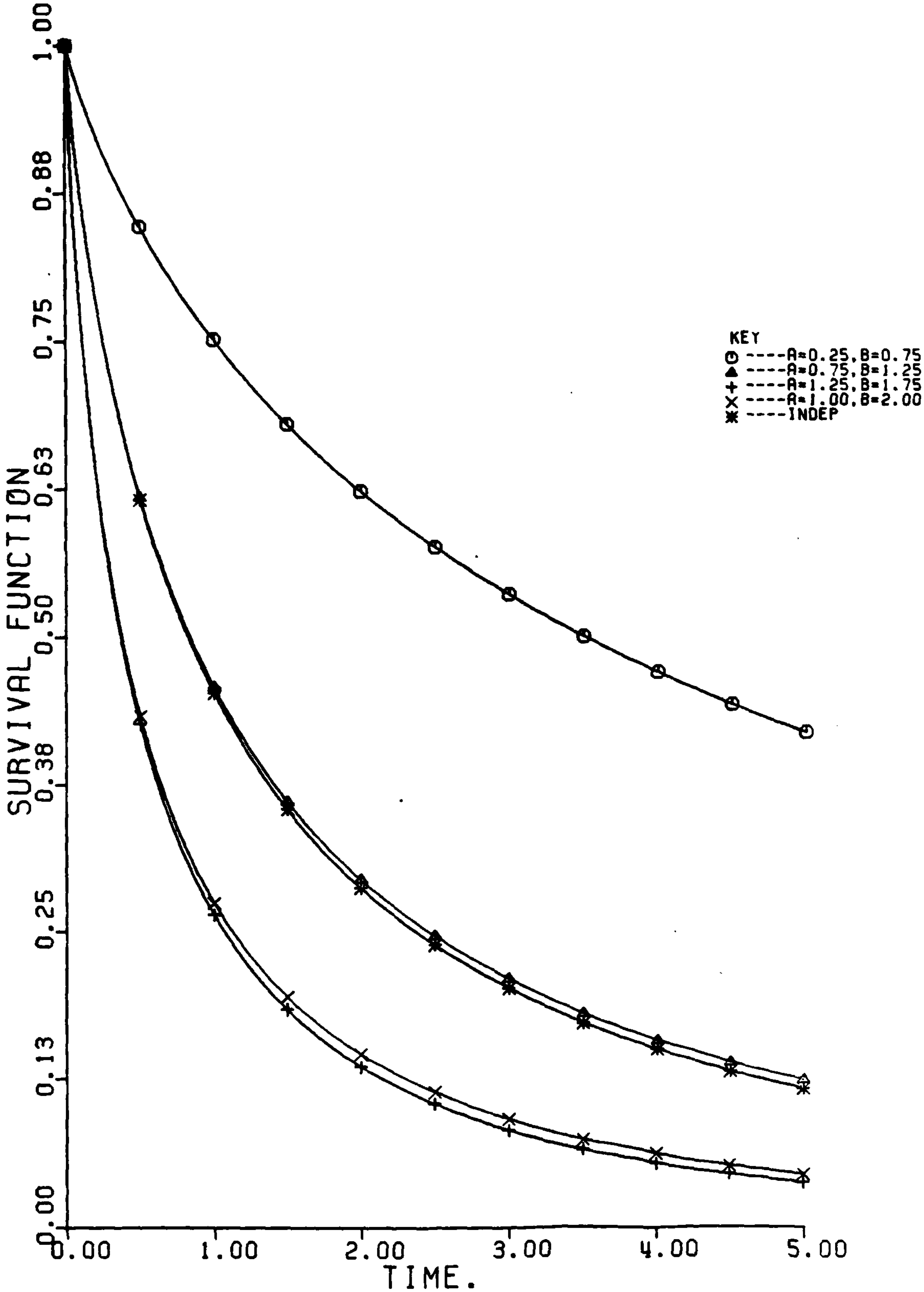
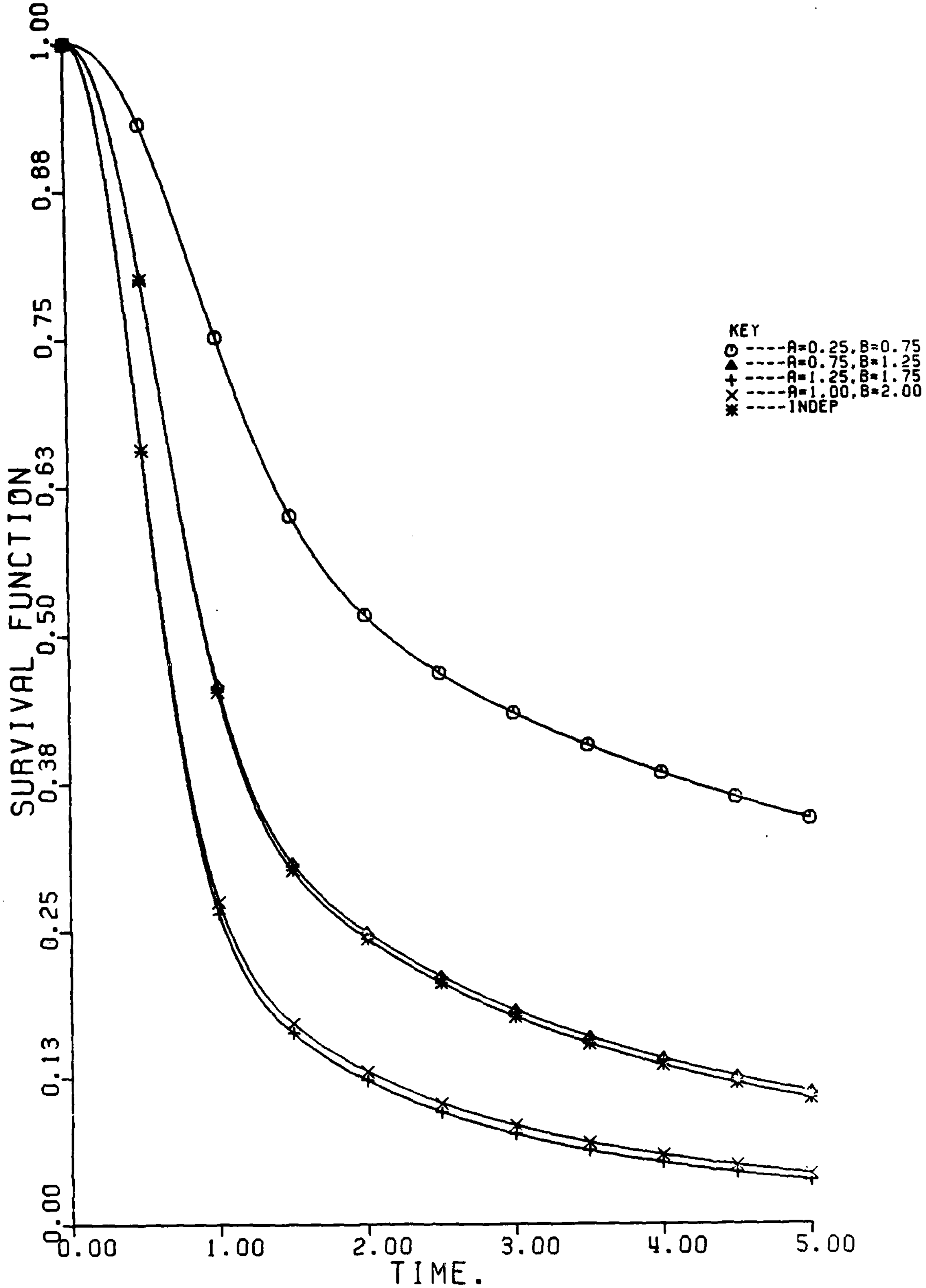


FIGURE 5 E

PARALLEL SYSTEM RELIABILITY UNDER UNIF (A,B) MODEL
FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=1/2, \quad \eta_2=2.0.$



η, not $\lambda_1, \lambda_2, \eta$, a, b. However, in many instances we have extensive data on the performance of the components in the lab under ideal operating conditions so that one may consider $\lambda_1, \lambda_2, \eta$ to be known based on estimates from this data. We shall focus on the problem of estimating θ and a, based on data on the system failure times collected in the operating environment. Let $t_1, \ldots, t_n$ be the failure times for n such systems put on test, and, let $w_i = t_i^{\eta}$, i = 1, ..., n.

Prior to attempting to estimate (a, θ), we would like to check if the model (3.1) is feasible. A graphical check of this model can be done through the scaled total time on test (STTOT) plot of Barlow and Campo (1975). The STTOT for W is

$$G_W(t) = \frac{\int_0^{F^{-1}(t)} Rs(t)dt}{\int_0^{F^{-1}(1)} Rs(t)dt} = 1-(1-t)^{(a-1)/a} \text{ for } a > 1. \tag{3.2}$$

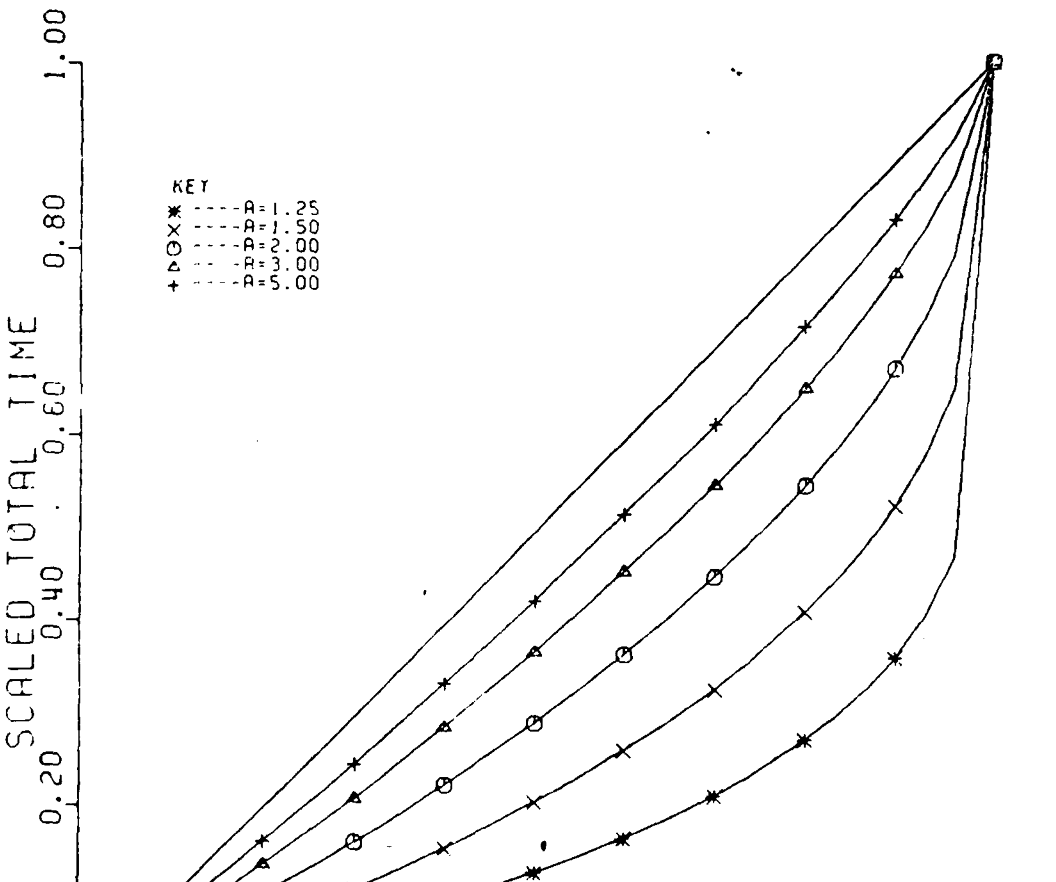
Note that (3.2) depends only on a. Figure 6 shows the form of the STTOT for several values of a. Notice that for all a, the STTOT is below the 45° line (which corresponds to exponential system life) since the hazard rate of the series system is decreasing. Let

$$Tn(W_{(i)}) = \sum_{j=1}^{i} W_{(j)} + (n-i)W_{(i)}, \tag{3.3}$$

where $W_{(1)} \leq W_{(2)} \leq \ldots \leq W_{(n)}$ are the ordered systems failure times be the total time on test at $W_{(i)}$. The empirical STTOT plot then plots $(i/n, T_n(W_{(i)})/T(W_{(n)}))$ which can be compared to figure 6 for a graphical check of the model. Also, crude estimates of a can be obtained by

FIGURE 6

SCALED TOTAL TIME ON TEST TRANSFORM
FOR GAMMA MODEL.



comparing the empirical and theoretical STTOT plots. When there is no random environmental effect and the components are independent, then the empirical STTOT plot should look like the 45° line. Also as a tends to infinity this plot approaches the 45° line.

We now consider several estimates of a and θ. The log likelihood for the model (3.1), based on a sample of size n, is

$$L(a,\theta) = n \ln a + n \ln \theta - (a+1)\sum_{i=1}^{n} \ln (1+\theta W_i) \tag{3.4}$$

so that

$$\partial/\partial a \; L(a,\theta) = n/a - \sum_{i=1}^{n} \ln (1 + \theta W_i) \tag{3.5}$$

and $$\partial/\partial\theta \; L(a,\theta) = n/\theta - (a+1) \sum_{i=1}^{n} w_i/(1+ \theta w_i) \tag{3.6}$$

For (3.5) we note that the maximum likelihood estimator of a

$$a_{mle} = \frac{n}{\sum_{i=1}^{n} \ln (1 + \theta W_i)} \tag{3.7}$$

and the maximum likelihood estimator of θ is the solution to

$$\frac{n}{\theta} - \left(\frac{n}{\sum \ln (1 + \theta W_i)} + 1\right) \left(\sum \frac{W_i}{1 + \theta W_i}\right) = 0. \tag{3.8}$$

One can show that θ is positive if $n \sum_{i=1}^{n} w_i^2 > 2(\sum_{i=1}^{n} w_i)^2$. (3.9)

In such case θ_{mle} is obtained by solving

(3.8) numerically.

A second estimator of (a, θ) is the method of moments (mme). Since $E(W) = [\theta(a-1)]^{-1}$ and $E(W^2) = 2[\theta^2(a-1)(a-2)]^{-1}$ where $a > 2$, we have

$$a_{mme} = 1 + \frac{\sum w_i^2}{\sum w_i^2 - 2(\sum w_i^2)} \quad (3.10) \text{ and } \theta_{mme} = \frac{\sum w_i^2 - 2(\sum w_i)^2}{\sum w_i(\sum w_i)^2} \qquad (3.11)$$

provided that (3.9) holds. If (3.9) does not hold, then this estimator does not exist.

A third estimator was suggested by Berger (1983) in a different context. He suggested estimating θ a modified methods of moments estimator $\theta_{ber} = (a\, w)^{-1}$, (3.12)

where $w = \sum w_i/n$,

which is used as the true value of θ in the likelihood (3.4) so that the estimator of a is the solution to

$$-\sum \ln\left(1 + \frac{w_i}{aw}\right) + \frac{(a+1)}{wa^2} \sum \frac{w_i}{1 + w_i/(aw)} = 0 \qquad (3.13)$$

A final estimator is based on the STTOT plot. Let $C_i = \ln(1 - i/n)$ and $D_i = \ln(1 - Tn(W_{(i)})/Tn(W_{(n)}))$, $i = 1, \ldots, n-1$. If (3.2) holds, then we should have

$$D_i = (1 - 1/a)C_i,\ i = 1, \ldots, n-1, \qquad (3.14)$$

so the value of a which minimizes

$\sum_{i=1}^{n-1} (D_i - (1-1/a)\, C_i)^2$ is a reasonable estimator of a

The resulting estimator is $a_{ls} = \dfrac{\sum C_i^2}{\sum C_i^2 - \sum C_i D_i}$ (3.15)

FIGURE 7 A

SCALED TOTAL TIME ON TEST PLOT
FOR SIMULATED DATA.

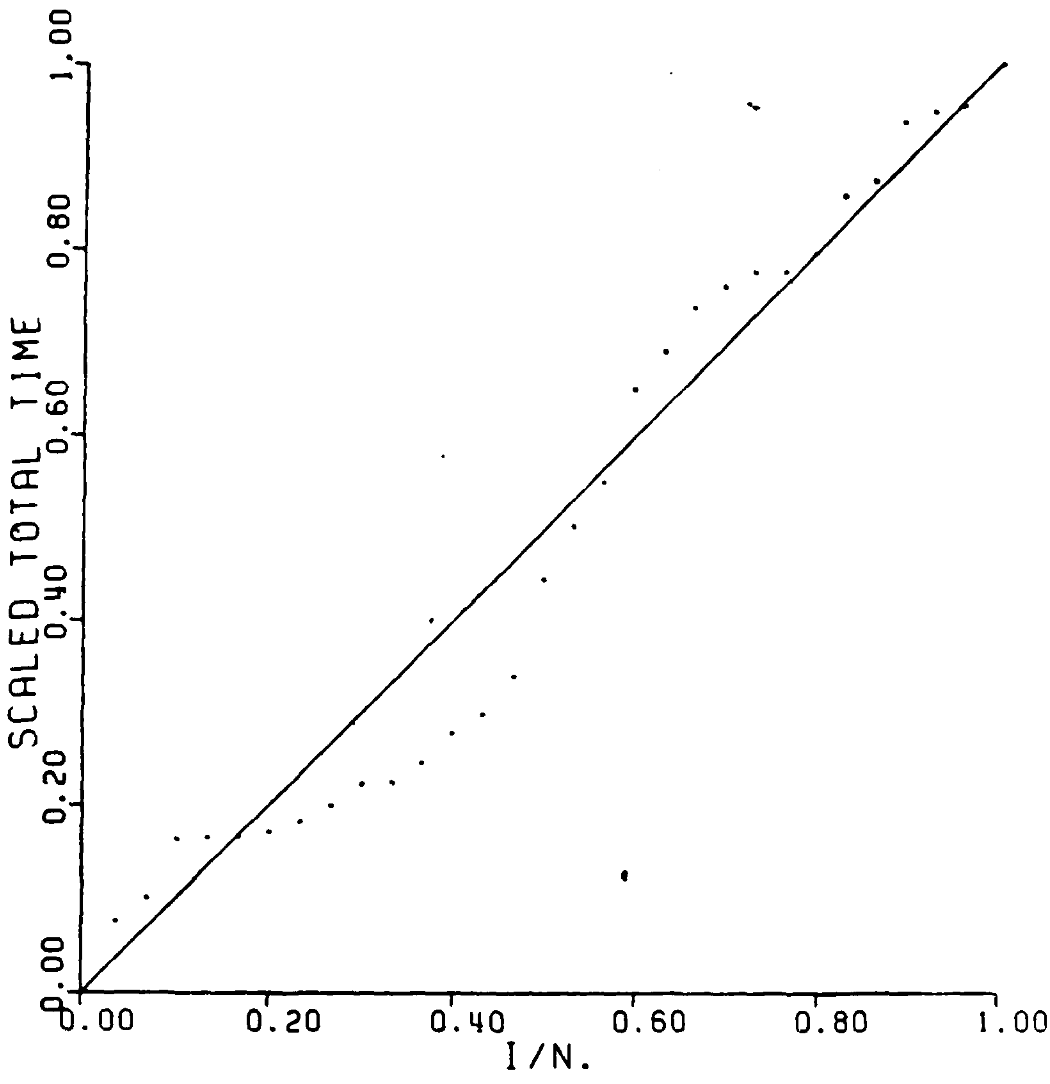
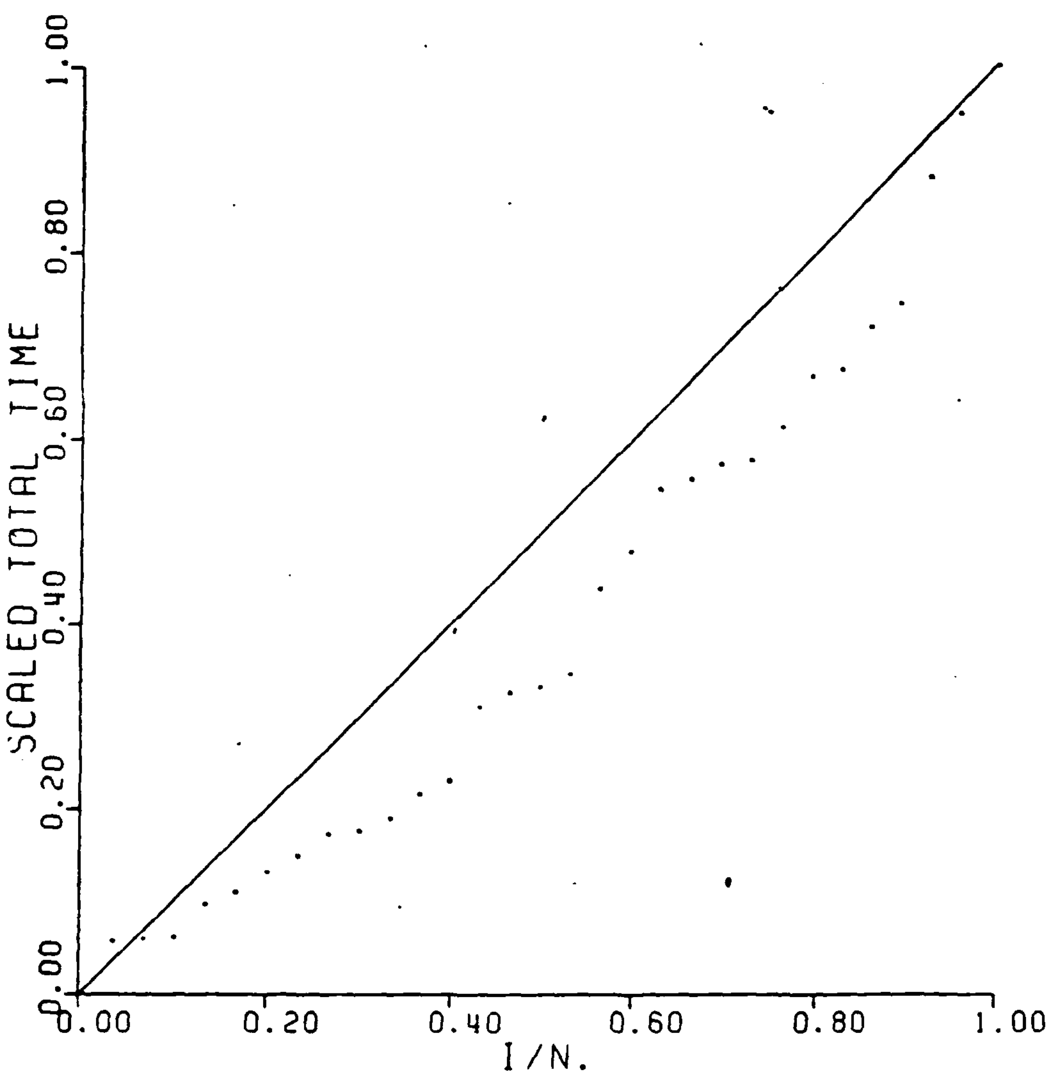


FIGURE 7 B

SCALED TOTAL TIME ON TEST PLOT
FOR SIMULATED DATA.



which is in the parameter space if $\sum C_i^2 > \sum C_i D_i$. A better estimator should be obtained by weighting the D_i's differently since for $i < j$, $\mathrm{Var}(D_i) < \mathrm{Var}(D_j)$. The variance of D_i depends on the unknown parameter a so we weight by the variance of D_i computed under an assumed exponential distribution. The variance of D_i in that case is

$$V_i = \sum_{j=1}^{i} \frac{1}{(n-j)^2}, \quad i = 1, \ldots, n-1 \tag{3.16}$$

so that the weighted least squares estimator of a is

$$a_{wls} = \frac{\sum C_i^2/V_i}{\left(\sum \frac{C_i^2}{V_i} - \sum \frac{C_i D_i}{V_i}\right)} \text{ if } \sum C_i^2/V_i > \sum C_i D_i/V_i. \tag{3.17}$$

Once we have obtained a by either of the two least squares estimators, we substitute this value into (3.6) and solve this equation numerically for θ_{ls} or θ_{wls}.

The condition $\sum C_i^2/V_i > \sum C_i D_i/V_i$ includes a few more possible samples than the condition (3.9) for the other three estimators. However, those samples which satisfy $\sum C_i^2/V_i > \sum C_i D_i /V_i$ for which (3.9) fails to obtain yield very large estimates of θ. Since a reasonable model for T when θ and a are not estimable is the independent Weibull series system which has system reliability very close to (3.1) when a is very large, this is not a problem. Figures 7a and 7b are scaled total time on test plots from two simuilated samples of size 30 from (3.1) with $a = 3$, $\theta = 1$. Looking at figure 7a, we see that the estimated scaled total on test doesn't look too different from the 45° so that an exponential model might not be unreasonable. For this data set only the weighted least squares estimator exists and it yields $a_{WLS} = 45.33$ and $\theta = .0567$. For the data in figure 7b all estimates exist, and we have

$\theta_{mle} = .93$ $a_{mle} = 2.98$

$\theta_{mme} = .491$ $a_{mme} = 4.86$

$\theta_{ber} = .720$ $a_{ber} = 7.02$

$\theta_{ls} = .739$ $a_{ls} = 3.58$

$\theta_{wls} = .970$ $a_{wls} = 2.89$

To study the properties of these estimators, a small scale Monte Carlo study was performed. Random samples of size n = 15, 30, 50, 75, or 100 were generated with $\lambda_1 + \lambda_2 = 3$, b = 3, so $\theta = 1$ and a = 2, 3, 5. 1000 samples were generated for each combination of n and a. The bias, standard deviation of the estimates and n, the number of samples where the estimator exists is reported in table 1 for a, table 2 for θ, and in table 3 for an estimator of the system reliability obtained from (3.1) at $t_0 = 9.085$. The true system reliability at t_0 is .8255 when a = 2, .75 when a = 3, and .619 when a = 5. Also reported in each table is the bias and standard deviation of the least square and weighted least square estimators when they are restricted to those samples where the other estimators exist.

From these tables we note that Berger's modified estimator performs very poorly. Also the weighted least squares estimator allows for estimation of parameters in many more samples when n is small. In general the maximum likelihood estimator outperforms the other estimators, however, when the weighted least squares estimator is restricted to those samples where the maximum likelihood estimator exists, this estimator performs much better when n is small. The somewhat better performance of the MLE in terms of bias is deceptive since some of the estimates of a are less than one, which implies that the mean system reiliability is infinite. Also the weighted least squares estimator of system reliability seems to outperform the other estimators of the system reliability in spite of its relatively poor performance as an estimator of θ. Our recommendation is to use the weighted least squares estimator since it more often provides estimators of the relevant parameters and is somewhat easier to compute.

TABLE 1

BIAS AND STANDARD DEVIATION(SD) OF ESTIMATORS OF A

		MAXIMUM LIKELIHOOD			WEIGHTED LEAST SQUARES			LEAST SQUARES			METHOD OF MOMENTS			BERGER'S METHOD		
K	N	M	BIAS	SD	M	BIAS	SD	M	BIAS	SD	M	BIAS	SD	M	BIAS	SD
2	15	769	4.5	29.	852	4.8	41.	762	6.8	49.	770	9.1	37.	770	14.4	65.
					766	1.3	5.	715	3.5	16.						
2	30	916	2.8	20.	953	4.7	37.	877	6.4	52.	916	6.1	32.	916	10.0	52.
					912	1.1	3.	857	5.4	51.						
2	50	979	5.8	114.	989	1.7	10.	956	4.0	16.	979	8.5	131.	979	15.4	241.
					976	1.0	3.	952	3.5	12.						
2	75	996	0.9	4.	998	1.0	5.	974	2.4	12.	996	2.2	4.	996	4.5	8.
					996	1.0	5.	972	2.4	12.						
2	100	999	0.5	3.	1000	1.7	35.	989	1.5	9.	999	1.7	5.	999	3.7	8.
					999	0.6	2.	989	1.5	9.						
3	15	64.	7.3	39.	753	36.4	843.	653	13.1	149.	643	16.8	77.	643	26.0	114.
					636	1.3	9.	573	10.7	159.						
3	30	809	5.7	30.	870	13.8	141.	768	11.9	100.	810	9.9	104.	809	17.7	68.
					804	1.8	6.	731	11.1	102.						
3	50	916	3.6	18.	935	6.9	65.	864	6.4	33.	916	6.6	25.	916	12.5	4[illegible].
					912	2.9	29.	851	5.7	32.						
3	75	963	2.5	14.	977	2.8	17.	925	11.6	144.	963	4.7	24.	963	9.6	38.
					958	1.3	5.	923	11.6	144.						
3	100	978	1.7	7.	989	2.1	12.	956	3.7	19.	978	3.0	9.	978	7.2	15.
					978	1.3	5.	952	3.6	19.						
5	15	520	38.7	573.	665	8.2	53.	558	30.3	493.	522	69.8	925.	522	112.9	1505.
					516	-0.7	5.	458	1.8	15.						
5	30	674	20.4	148.	752	9.3	68.	669	13.7	109.	674	31.7	202.	674	56.[illegible]	347.
					660	3.2	29.	601	8.4	86.						
5	50	801	7.6	39.	850	9.0	97.	756	13.4	88.	801	11.4	52.	801	23.4	91.
					787	2.0	10.	722	8.6	56.						
5	75	893	12.8	139.	915	8.0	94.	827	6.6	22.	89.	15.3	122.	893	32.6	260.
					878	2.9	16.	714	5.8	20.						
5	100	89.	9.5	84.	913	19.6	307.	835	11.0	81.	89.	13.0	120.	89.	[illegible]	[illegible]
					879	13.7	273.	821	9.5	79.						

TABLE 2

BIAS AND STANDARD DEVIATION(SD) OF ESTIMATORS OF θ

		MAXIMUM LIKELIHOOD			WEIGHTED LEAST SQUARES			LEAST SQUARES			METHOD OF MOMENTS			BERGER S METHOD		
A	N	M	BIAS	SD	M	BIAS	SD	M	BIAS	SD	M	BIAS	SD	M	BIAS	SD
2	15	769	0.356	1.702	852	-.102	0.742	782	-.192	0.729	770	-.683	0.205	770	-.803	0.122
					766	-.027	0.025	715	-.156	0.701						
2	30	916	0.112	0.919	953	-.135	0.580	877	-.254	0.586	916	-.623	0.192	916	-.798	0.095
					912	-.100	0.567	857	-.239	0.584						
2	50	979	0.016	0.648	989	-.126	0.492	956	-.263	0.514	979	-.575	0.184	979	-.792	0.079
					976	-.115	0.486	952	-.260	0.513						
2	75	996	-.025	0.522	998	-.125	0.432	974	-.247	0.475	996	-.541	0.174	996	-.790	0.065
					996	-.124	0.431	972	-.246	0.474						
2	100	999	-.019	0.437	1000	-.101	0.381	989	-.216	0.423	999	-.508	0.153	999	-.785	0.052
					1000	-.101	0.381	989	-.216	0.423						
3	15	642	0.691	1.900	753	0.210	1.049	653	0.119	1.031	683	-.513	0.348	643	-.705	0.203
					636	0.390	1.040	573	0.245	1.038						
3	30	809	0.175	1.049	870	0.000	0.757	769	-.096	0.745	810	-.469	0.338	809	-.725	0.160
					804	0.074	0.740	731	-.057	0.743						
3	50	916	0.075	0.766	935	-.012	0.618	864	-.112	0.663	916	-.404	0.333	916	-.718	0.135
					912	0.011	0.609	851	-.100	0.660						
3	75	963	0.030	0.624	977	-.027	0.555	925	-.144	0.603	963	-.375	0.322	963	-.717	0.120
					958	-.010	0.546	923	-.142	0.603						
3	100	978	-.028	0.515	989	-.075	0.472	956	-.165	0.511	978	-.345	0.297	978	-.716	0.104
					978	-.055	0.465	952	-.161	0.509						
5	15	522	1.352	3.109	665	0.715	1.609	558	0.578	1.536	522	-.238	0.601	522	-.546	0.348
					516	1.084	1.599	458	0.827	1.561						
5	30	674	0.558	1.609	752	0.366	1.118	669	0.180	1.026	674	-.199	0.576	674	-.584	0.282
					660	0.523	1.104	601	0.286	1.029						
5	50	801	0.256	1.559	850	0.184	0.869	756	0.105	0.863	801	-.193	0.541	801	[illegible]	0.232
					787	0.267	0.850	722	0.150	0.858						
5	75	893	0.129	0.817	915	0.112	0.728	827	0.014	0.747	893	.189	[illegible]	[illegible]	-.628	[illegible]
					878	0.153	0.715	814	0.028	0.744						
5	100	892	0.033	0.683	913	0.020	0.628	835	-.064	0.666	892	.206	0.494	892	-.644	.135
					879	0.055	0.615	821	-.050	0.663						

TABLE 3

BIAS AND STANDARD DEVIATION(SD) OF ESTIMATORS OF SYSTEM RELIABILITY AT T=.9085

		MAXIMUM LIKELIHOOD			WEIGHTED LEAST SQUARES			LEAST SQUARES			METHOD OF MOMENTS			BERGER'S METHOD		
R	N	M	BIAS	SD	M	BIAS	SD	M	BIAS	SD	M	BIAS	SD	M	BIAS	SD
2	15	769	-.012	.0647	852	-.004	.0586	762	0.002	.0588	770	0.037	.0503	770	0.064	.0463
					766	-.006	.0577	715	0.002	.0586						
2	30	916	-.005	.0473	953	0.002	.0424	877	0.010	.0434	916	0.037	.0357	916	0.069	.0335
					912	0.001	.0426	857	0.009	.0434						
2	50	979	-.001	.0372	989	0.003	.0349	956	0.010	.0359	979	0.035	.0300	979	0.071	.0233
					976	0.003	.0348	952	0.010	.0359						
2	75	996	0.000	.0290	998	0.004	.0274	974	0.010	.0292	996	0.034	.0244	996	0.072	.0238
					996	0.003	.0274	972	0.010	.0293						
2	100	999	0.001	.0243	1000	0.004	.0234	989	0.010	.0248	999	0.034	.0222	999	0.075	.0216
					999	0.004	.0233	989	0.010	.0248						
3	15	642	-.018	.0815	753	-.010	.0767	653	-.006	.0748	643	0.031	.0661	642	0.067	.0616
					636	-.015	.0764	573	-.008	.0751						
3	30	809	-.007	.0577	870	-.003	.0552	769	0.002	.0551	810	0.024	.0490	809	0.062	.0472
					804	-.006	.0544	731	0.001	.0550						
3	50	916	-.003	.0429	935	-.001	.0412	864	0.005	.0431	916	0.022	.0366	916	0.066	.0340
					912	-.002	.0411	851	0.005	.0432						
3	75	963	-.001	.0372	977	0.000	.0356	925	0.007	.0374	963	0.021	.0327	963	0.067	.0313
					958	0.000	.0357	923	0.007	.0375						
3	100	9788	0.001	.0309	989	0.002	.0299	956	0.008	.0311	978	0.019	.0267	978	0.067	.0261
					978	0.002	.0299	952	0.008	.0312						
5	15	520	-.029	.1011	665	-.024	.0967	558	-.020	.0977	522	0.012	.0926	522	0.054	.0921
					516	-.030	.0968	458	-.022	.0968						
5	30	674	-.022	.0691	752	-.020	.0674	669	-.013	.0651	674	0.001	.0613	674	0.045	.0608
					660	-.027	.0665	601	-.016	.0652						
5	50	801	-.006	.0545	850	-.006	.0530	756	-.002	.0532	801	0.010	.0494	801	0.055	.0485
					787	-.008	.0528	722	-.002	.0535						
5	75	893	-.005	.0442	915	-.005	.0436	827	-.002	.0431	893	0.007	.0406	893	0.051	.0380
					878	-.006	.0430	814	-.002	.0432						
5	100	892	-.002	.0375	913	-.001	.0392	835	0.003	.0380	892	0.008	.0357	892	0.052	.0345
					879	-.002	.0390	821	0.003	.0381						

Acknowledgment

This work was supported by the U.S. Air Force Office of Scientific Research under contract AFOSR-82-0307.

Bibliography

Bain, L. J. (1978). Statistical Analysis of Reliability and Life-Testing Models. New York: Marcel Dekker, Inc.

Barlow, R. E. and Campo, R. (1975). Total Time on Test Processes and Applications to Failure Data Analysis. Reliability and Fault Tree Analysis. (Barlow, Fussell, and Singpurwalla, eds.). SIAM, Philadelphia, 451-481.

Basu, A. P. and Klein, J. P. (1982). Some Recent Results in Competing Risk Theory. Survival Analysis (Crowley and Johnson, eds.). The Institute of Mathematical Statistics, 216-227.

Clayton, D. G. (1978). A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familiar Tendency in Chronic Disease Incidence. Biometrica 65, 141-151.

Hui, S. L. and Berger, J. (1983). Empirical Bayes Estimation of Rates. Journal of the American Statistical Association 78, 753-760.

Hutchinson, T. P. (1981). Compound Gamma Bivariate Distributions. Metrika 28, 263-271.

Klein, J. P. and Moeschberger, M. L. (1984). Consequences of Departure from Independence in Exponential Series Systems. Technometrics 26, 277-284.

Lindley D. V. and Singpurwalla, N. D. (1985). Multivariate Distributions for the Reliability of a System of Components Sharing a Common Environment. George Washington University Technical Report.

Oakes, D. (1982). A Model for Association in Bivariate Survival Data. Journal of the Royal Statistical Society B, 44, 414-422.

Takahasi, K. (1965). Note on the multivariate Burr's Distribution. Annals of the Institute of Statistical Mathematics, Tokyo, 17, 257-260.

APPENDIX L

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY NA	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE NA	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION The Ohio State University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NM
6c. ADDRESS (City, State, and ZIP Code) 1314 Kinnear Road Columbus, Ohio 43212	7b. ADDRESS (City, State, and ZIP Code) AFOSR/NM Bldg 410 Bolling AFB DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-82-0307
8c. ADDRESS (City, State, and ZIP Code) AFOSR/NM Bldg 410 Bolling AFB DC 20332-6448	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61102F; PROJECT NO. 2304; TASK NO.; WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification) Estimating Reliability for Bivariate Exponential Distributions.

12. PERSONAL AUTHOR(S)
John P. Klein and Asit P. Basu

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 9-1-82 TO 12-31-87	May 31, 1988	8

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Minimum variance unbiased estimators; bivariate exponential; reliability; maximum likelihood estimator; jackknife; survival function
XXXX	XXXXXXXXXX	XXX	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The problem of estimating reliability for the bivariate exponential distributions of Block and Basu (1974) and Marshall and Olkin (1967) is considered. For Bock and Basu's model a minimum variance unbiased estimator of the joint survival function is obtained in the case of identically distributed marginals. For the non-identically distributed case the performance of the maximum likelihood estimator and the jackknifed maximum likelihood estimator is studied. For Marshall and Olkin's model the performance of several different parameter estimators and bias reduction techniques for estimating joint reliability are considered.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL Maj. Brian Woodruff	22b. TELEPHONE (Include Area Code) (202) 767-5027	22c. OFFICE SYMBOL NM

DD FORM 1473, 84 MAR   83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

Reprinted from

SANKHYĀ

THE INDIAN JOURNAL OF STATISTICS

SERIES B, VOLUME 47, PART 3, 1985

ESTIMATING RELIABILITY FOR BIVARIATE EXPONENTIAL DISTRIBUTIONS

By

JOHN P. KLEIN

and

ASIT P. BASU

STATISTICAL PUBLISHING SOCIETY

CALCUTTA

Sankhyā : The Indian Journal of Statistics
1985, Volume 47, Series B, Pt. 3, pp. 346-353.

ESTIMATING RELIABILITY FOR BIVARIATE EXPONENTIAL DISTRIBUTIONS

By JOHN P. KLEIN*

The Ohio State University, Columbus

and

ASIT P. BASU**

University of Missouri, Columbia

SUMMARY. The problem of estimating reliability for the bivariate exponential distributions of Block and Basu (1974) and Marshall and Olkin (1967) is considered. For Book and Basu's model a minimum variance unbiased estimator of the joint survival function is obtained in the case of identically distributed marginals. For the non-identically distributed case the performance of the maximum likelihood estimator and the jackknifed maximum likelihood estimator is studied. For Marshall and Olkin's model the performance of several different parameter estimators and bias reduction techniques for estimating joint reliability are considered.

1. Introduction

Let X, Y have either the bivariate exponential distribution (BVE) of Marshall and Olkin (1967) or the absolutely continuous bivariate exponential distribution (ACBVE) of Block and Basu (1974). These two distributions have found considerable use as models for both physical and biological systems. The problem of interest is to estimate the joint reliability function, $\bar{F}(x, y) = P(X > x, Y > y)$, for each of these two distributions. A natural estimator of $\bar{F}(x, y)$ is obtained by substituting in the appropriate expression for $\bar{F}(x, y)$ good estimators of the model parameters. Often, as seen in Pugh (1963), Basu (1964) or Basu and El Mawaziny (1978), such estimators can be considerably biased. We wish to obtain reduced biased estimators of $\bar{F}(x, y)$ for both the BVE and ACBVE distributions.

In Section 2 this estimation problem is considered for the ACBVE. In the case of identically distributed marginals, using the Rao-Blackwell and the

AMS (1980) *subject classification :* 62N05

Key words : Minimum variance unbiased estimators ; Bivariate exponential ; Reliability ; Maximum likelihood estimator ; Jackknife ; Survival function.

*Research was supported in part by contract AFOSR-82-0307 for the Air Force office of Scientific Research.

**Research was supported in part under contract N00014-78-C-0655 for the office of Naval Research.

Lehmann-Scheffé theorems we obtain minimum variance unbiased estimators (UMVUE) of $\bar{F}(x, y)$. In the case of non-identically distributed marginals this approach fails since there are no complete sufficient statistics. Here we investigate the performance of the maximum likelihood estimator as well as the jackknifed maximum likelihood estimator.

In Section 3 we consider the estimation of $\bar{F}(x, y)$ for the BVE. Again there are no complete sufficient statistics so no minimum variance unbiased estimators can be obtained. Several different methods for estimating parameters are considered. For each estimation procedure we consider several bias reduction techniques.

2. Absolutely continuous bivariate exponential

2.1 *Introduction*. Let (X, Y) have the absolutely continuous bivariate exponential distribution of Block and Basu (1974) with parameters $\lambda_1, \lambda_2 > 0$, $\lambda_{12} > 0((X, Y) \sim \text{ACBVE}(\lambda_1, \lambda_2, \lambda_{12}))$. This distribution is closely related to the bivariate exponential of Freund (1961). It has been used by Gross, Clark and Lui (1971) and Gross (1973) to model the lifetimes of two organ systems and by Gross and Lam (1981) for modeling paired survival time data such as survival of a tumor remission when a patient receives two types of treatment.

For this model the joint reliability function is

$$\bar{F}(x, y) = \frac{\lambda}{(\lambda_1+\lambda_2)} \exp(-\lambda_1 x - \lambda_2 y - \lambda_{12} \max(x, y))$$

$$- \frac{\lambda_{12}}{(\lambda_1+\lambda_2)} \exp(-\lambda \max(x, y)), \text{ for } x, y > 0, \quad \dots \quad (2.1.1)$$

with $\lambda = \lambda_1+\lambda_2+\lambda_{12}$.

This distribution has the bivariate loss of memory property (LMP) defined by Block and Basu (1974). It is the absolutely continuous part of the Marshall and Olkin (1967) bivariate exponential.

We shall consider two cases for estimating $\bar{F}(x, y)$, one where the marginals are identically distributed and the general model.

2.2 *Equal marginals*. Consider the model (2.1.1) with $\lambda_1 = \lambda_2 = \alpha$ and $\lambda_{12} = \beta$. Let $(x_1, y_1), \dots, (x_n, y_n)$ be a random sample from (2.1.1). Let $U_1 = \Sigma \max(x_i, y_i)$ and $U_2 = \Sigma(x_i+y_i)$. Mehrotra and Michalek (1976) show that (U_1, U_2) is a complete sufficient statistic for (α, β). The MLE of α, β are given by

$$\hat{\alpha} = n\left(\frac{1}{u_2-u_1} - \frac{1}{2u_1-u_2}\right), \hat{\beta} = n\left(\frac{2}{2u_1-u_2} - \frac{1}{u_2-u_1}\right) \quad \ldots \quad (2.2.1)$$

These estimators are biased by a factor of $n/(n-1)$ so the estimators $\tilde{\alpha} = \frac{n-1}{n}\hat{\alpha}$ and $\tilde{\beta} = \frac{n-1}{n}\hat{\beta}$ are the UMVUE of α and β. Two natural estimators of $\bar{F}(x, y)$ are obtained by sustituting either $(\hat{\alpha}, \hat{\beta})$ or $(\tilde{\alpha}, \tilde{\beta})$ in (2.1.1).

We now use the method proposed by Basu (1964) to obtain the UMVUE of $\bar{F}(x, y)$.

Define

$$\phi(x, y : X, Y) = \begin{cases} 1 & \text{if } X > x,\ Y > y \\ 0 & \text{otherwise.} \end{cases} \quad \ldots \quad (2.2.2)$$

Clearly $\phi(x, y : X, Y)$ is an unbiased estimator of $\bar{F}(x, y)$ based on a random sample of size one from a ACBVE (α, α, β). By the Rao-Blackwell and Lehmann-Scheffé theorems the estimator $\tilde{\bar{F}}(x, y) = E(\phi(x, y; X, Y) \mid u_1, u_2)$ is the UMVUE of $\bar{F}(x, y)$.

To simplify the calculations let $T = U_2 - U_1$ and $V = 2U_1 - U_2$, that is $T = \Sigma \min(X_i, Y_i)$ and $V = \Sigma \max(X_i, Y_i) - \Sigma \min(X_i, Y_i)$. From Mehrotra and Michalek (1976), the joining density of (T, V) is

$$f(t, v) = \begin{cases} \dfrac{(2\alpha+\beta)^n(\alpha+\beta)^n}{[(n-1)!]^2} t^{n-1} v^{n-1} \exp(-(2\alpha+\beta)t - (\alpha+\beta)v), & t, v > 0 \\ 0 & \text{otherwise.} \end{cases} \quad \ldots \quad (2.2.3)$$

Now split the sample of size n into two independent subsamples of sizes one and $n-1$, respectively. Let (Z_1, Z_2) denote the sample of size one and let T_1, V_1 denote the statistics T and V defined on the remaining $n-1$ observations. The joint density of (Z_1, Z_2, T_1, V_1) is

$$f(z_1, z_2, t_1, v_1) = \begin{cases} \dfrac{(2\alpha+\beta)^n(\alpha+\beta)^n}{2[(n-2)!]^2} t_1^{n-2} v_1^{n-2} \exp[-(2\alpha+\beta)t_1 - (\alpha+\beta)v_1 \\ \qquad -\alpha z_1 - (\alpha+\beta)z_2] \quad \text{if } z_1 < z_2 \\ \dfrac{(2\alpha+\beta)^n(\alpha+\beta)^n}{2[(n-2)!]^2} t_1^{n-2} v_1^{n-2} \exp[-(2\alpha+\beta)t_1 - (\alpha+\beta)v_1 \\ \qquad -(\alpha+\beta)z_1 - \alpha z_2] \quad \text{if } z_2 < z_1 \end{cases} \quad \ldots \quad (2.2.4)$$

for $t_1, v_1 > 0$.

Clearly $V = V_1+\max(Z_1, Z_2)-\min(Z_1, Z_2)$ and $T = T_1+\min(Z_1, Z_2)$. Hence the joint density of (Z_1, Z_2, T, V) is

$$
\begin{aligned}
f(z_1, z_2, t, v) &= \frac{(2\alpha+\beta)^n(\alpha+\beta)^n}{2[(n-1)\,!]^2}(t-z_1)^{n-2}(v-z_2+z_1)^{n-2}\exp(-(2\alpha+\beta)t-(\alpha+\beta)v) \\
&\qquad \text{for } 0 < v, t;\ 0 < z_1 < t;\ z_1 < z_2 < v+z_1 \\
&= \frac{(2\alpha+\beta)^n(\alpha+\beta)^n}{2[(n-1)\,!]^2}(t-z_2)^{n-2}(v-z_1+z_2)\exp(-(2\alpha+\beta)t-(\alpha+\beta)v) \\
&\qquad \text{for } 0 < v, t;\ 0 < z_2 < t;\ z_2 < z_1 < v+z_2. \quad \ldots \quad (2.2.5)
\end{aligned}
$$

Thus the conditional distribution of Z_1, Z_2 given T, V is

$$
f(z_1, z_2 \mid T=t,\ V=v)= \begin{cases} \dfrac{(n-1)^2(t-z_1)^{n-2}(v-z_2+z_1)^{n-2}}{2t^{n-1}v^{n-1}}, & 0 < z_1 < t, \\ & z_1 < z_2 < v+z_1 \\ \dfrac{(n-1)^2(t-z_2)^{n-2}(v-z_1+z_2)^{n-2}}{2t^{n-1}v^{n-1}}, & 0 < z_2 < t, \\ & z_2 < z_1 < v+z_2. \end{cases} \quad \ldots \quad (2.2.6)
$$

To find $E(\phi(x, y; Z_1, Z_2) \mid T = t,\ V = v)$ consider three cases:—

Case 1 : $t > x = y > 0$. Here,

$$
\begin{aligned}
E(\phi(x, y; Z_1, Z_2) &\mid T = t,\ V = v) \\
&= 2\int_x^t\int_{z_1}^{v+z_1}\frac{(n-1)^2(t-z_1)^{n-2}(v-z_2+z_1)^{n-2}}{2t^{n-1}v^{n-1}}\,dz_2dz_1 = \left(1-\frac{x}{t}\right)^{n-1}. \quad \ldots \quad (2.2.7)
\end{aligned}
$$

Case 2 : $x < y < t$. Here,

$$
\begin{aligned}
E(\phi(x, y; &Z_1, Z_2) \mid T = t,\ V = v) \\
&= \int\int_{\{(z_1, z_2)\,:\,z_1 > y,\ z_2 > y\}} f(z_1, z_2 \mid t, v)dz_1dz_2 + \int\int_{\{(z_1, z_2)\,:\,x < z_1 < y,\ y < z_2\}} f(z_1, z_2 \mid t, v)dz_1dz_2 \\
&= \left(1-\frac{y}{t}\right)^{n-1} + \int_x^y\int_y^{v+z_1}\frac{1}{2}\,\frac{(n-1)^2(t-z_1)^{n-2}(v+z_1-z_2)^{n-2}}{t^{n-1}v^{n-1}}\,dz_2dz_1 \\
&= \left(1-\frac{y}{t}\right)^{n-1} + \frac{(n-1)}{2v^{n-1}t^{n-1}}\sum_{k=0}^{n-1}(-1)^k\frac{\binom{n-1}{k}(v-y+t)^{n-1-k}}{(n+k-1)} \\
&\qquad [(t-x)^{n+k-1}-(t-y)^{n+k-1}]. \quad \ldots \quad (2.2.8)
\end{aligned}
$$

Case 3 : $t > x > y > 0$. By symmetry

$E(\phi(x, y; Z_1, Z_2) | T = t, V = v)$

$$= \left(1-\frac{x}{t}\right)^{n-1} + \frac{(n-1)}{2v^{n-1}t^{n-1}} \sum_{k=0}^{n-1} \frac{(-1)^k \binom{n-1}{k}}{(n+k-1)} (v-x+t)^{n-1-k}$$

$$[(t-y)^{n-k-1}-(t-x)^{n+k-1}]. \qquad \ldots \quad (2.2.9)$$

2.3 *Unequal marginals.* When (X, Y) is ACBVE $(\lambda_1, \lambda_2, \lambda_{12})$ with λ_1 not known to be equal to λ_2 there does not exist a set of complete, sufficient statistics for $(\lambda_1, \lambda_2, \lambda_{12})$. Hence, the technique described in Section 2.2 fails. Maximum likelihood estimators of $(\lambda_1, \lambda_2, \lambda_{12})$ are obtained numerically by maximizing the likelihood function as described in Block and Basu (1974). The maximum likelihood estimator, $\hat{\bar{F}}(x, y)$ of $\bar{F}(x, y)$ is obtained by substituting these estimators into (2.1.1).

For small sample sizes, this estimator may be highly biased. To reduce this bias we consider the jackknifed version of the MLE estimator. This estimator is constructed as follows : Let $\hat{\bar{F}}^{(j)}_{n-1}(x, y)$ be the MLE of $\bar{F}(x, y)$ based on the subsample of size $n-1$ obtained by deleting the j-th observation from the original sample. The jackknifed version of $\hat{\bar{F}}(x, y)$ is

$$\hat{\bar{F}}_{jack}(x, y) = n\hat{\bar{F}}(x, y) - \frac{(n-1)}{n} \sum_{j=1}^{n} \hat{\bar{F}}^{j}_{n-1}(x, y). \qquad \ldots \quad (2.3.1)$$

Miller (1974) shows that this estimator removes the n^{-1}-th order term in the expansion of the bias of $\bar{F}(x, y)$.

To study the performance of the MLE and the jackknifed MLE of $\bar{F}(x, y)$, a simulation study was performed. For various values of $\lambda_1, \lambda_2, \lambda_{12}$ and n, 500 ACBVE samples were generated by the method of Friday and Patil (1977). Values of (x, y) were picked so that $\bar{F}(x, y) = \cdot 9$. The study showed that the jackknifed maximum likelihood estimator had significantly smaller bias than the MLE. For sample sizes of 10 or larger, the bias of this estimator is not statistically different from zero. However, the jackknifed MLE has a slightly larger mean squared error than the MLE in all cases considered.

3. Bivariate exponential

3.1 *Parameter estimation.* We say (X, Y) follows the bivariate exponential distribution of Marshall and Olkin (1967) with parameters $\lambda_1 > 0$, $\lambda_2 > 0$, and $\lambda_{12} \geqslant 0((X, Y)$ is BVE $(\lambda_1, \lambda_2, \lambda_{12}))$ if the joint survival function is

$$P(X > x, Y > y) = \bar{F}(x, y) = \exp(-\lambda_1 x - \lambda_2 y - \lambda_{12} \max(x, y)) \qquad \ldots \quad (3.1.1)$$

for $x > y > 0$. This distribution is not absolutely continuous since $P(X = y) = \lambda_{12}/(\lambda_1+\lambda_2+\lambda_{12})$. The marginals are exponential as is $\min(X, Y)$. This is the only bivariate distribution with exponential marginals and the loss of memory property (LMP) as defined in Block and Basu (1974).

To estimate $\lambda_1, \lambda_2, \lambda_{12}$ based on a random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$, let n_1, n_2, n_{12} be the number of observations with X_i less, greater, and equal to Y_i, respectively. Let $T = \Sigma \max(X_i, Y_i)$, $S_x = \Sigma X_i$, $S_y = \Sigma Y_i$. Bhattacharyya and Johnson (1971) show that (n_1, n_2, t, s_x, s_y) are jointly minimal sufficient but not complete. Hence, the approach of Section 2.1 cannot be applied. The maximum likelihood estimators are obtained by numerically maximizing the likelihood equations. Bhattacharyya and Johnson (1971) obtain conditions under which the MLE exist, and show that these estimators are asymptotically trivariate normal with mean $(\lambda_1, \lambda_2, \lambda_{12})$.

Bemin, Bain and Higgins (1972) have obtained method of moments estimators of the parameters. Proschan and Sullo (1976) obtained estimators of the parameters by using a first iterate in the likelihood equations. Arnold (1968) gives estimators of λ_i based on n_1, n_2, n_{12} and $U = \Sigma \min(X_i, Y_i)$. In the competing risks framework where only the minimum of X and Y is observed, these estimators are the unique minimum variance unbiased estimators of λ_i. All of the above estimators are asymptotically trivariate normal with mean $(\lambda_1, \lambda_2, \lambda_{12})$.

3.2 *Estimation of tail probability*. The problem of interest is to estimate $\bar{F}(x, y)$ given by (3.1.1). A natural method of estimating (3.1.1) is to use one of the above methods to estimate $(\lambda_1, \lambda_2, \lambda_{12})$ and substitute these estimates in (3.1.1).

Several methods may be used to reduce the bias of these estimators. The first approach is to expand the substitution estimator in a Taylor series about $(\lambda_1, \lambda_2, \lambda_{12})$ keeping only second order terms. When $E(\hat{\lambda}_i) = \lambda_i$, the bias of the substitution estimator is approximately equal to

$$E(\hat{\bar{F}}_{SUB}(x, y)) \cong \bar{F}(x, y)[1+\sigma^2/2]$$

where

$$\sigma^2 = (x, y, \max(x, y))\Sigma(x, y, \max(x, y))' \quad \ldots \quad (3.2.1)$$

and Σ is the appropriate covariance matrix of $(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_{12})$. This suggests a reduced bias estimator of $\bar{F}(x, y)$ given by

$$\hat{\bar{F}}_{TS}(x, y) = \hat{\bar{F}}_{SUB}(x, y)/[1+\hat{\sigma}^2/2] \quad \ldots \quad (3.2.2)$$

where $\hat{\sigma}^2$ is an estimator of σ^2.

A second approach to the bias of $\hat{\hat{\bar{F}}}_{SUB}(x, y)$ is through asymptotic theory. Note that $\ln \hat{\hat{\bar{F}}}_{SUB}(x, y)$ is asymptotically normal with mean $-\lambda_1 x-\lambda_2 y-\lambda_{12}\max(x, y)$ and variance σ^2. Thus, for large n, $\hat{\hat{\bar{F}}}_{SUB}(x, y)$ has a log normal distribution and

$$E(\hat{\hat{\bar{F}}}_{SUB}(x, y)) \to \bar{F}(x, y)\exp(\sigma^2/2) \qquad \dots \quad (3.2.3)$$

and $V(\hat{\hat{\bar{F}}}_{SUB}(x, y)) \to \bar{F}(x, y))^2 e^{\sigma^2}(e^{\sigma^2}-1)$. This suggests a reduced bias estimator of $\bar{F}(x, y)$ given by

$$\hat{\hat{\bar{F}}}_{LN}(x, y) = \hat{\hat{\bar{F}}}_{SUB}(x, y)\exp(\hat{\sigma}^2/2). \qquad \dots \quad (3.2.4)$$

A third method to reduce the bias of $\bar{F}_{SUB}(x, y)$ is the jackknife as described in Section 2.3.

To compare these estimators, a simulation study was performed. 500 BVE observations were generated for various combinations of n, $\lambda_1, \lambda_2, \lambda_{12}$. Values of (x, y) were selected so that $\bar{F}(x, y) = \cdot 9$.

Several conclusions can be drawn from the study. First, for all bias reduction techniques, those based on Arnold's estimators have a significantly larger mean squared error but a smaller relative bias. Secondly, there appears to be very little difference in the estimators based on the other three methods. For Arnold's estimators, all three bias reduction techniques yield approximately unbiased estimators with comparable mean squared errors. For the other methods, only the jackknifed estimator is approximately unbiased due to bias of the estimators of the parameters themselves since first order (bias) terms were neglected in the derivation of (3.2.2) and (3.2.4). The expansions based on Arnold's estimators are correct since these estimators of λ_1, λ_2 and n_{12} are unbiased. Our recommendation is to jackknife either the Proschan and Sullo estimator or the method of moments estimator since these are computationally easier than the MLE and have the smallest bias and mean square error of the three bias reduction methods.

References

Arnold, B. C. (1968): Parameter estimation for a multivariate exponential distribution, *Jour. Amer. Statist. Assoc.*, **63**, 848-52.

Basu, A. P. (1964): Estimates of reliability for some distributions useful in life-testing, *Technometrics*, **6**, 215-219.

Basu, A. P. and El Mawaziny (1978): Estimates of reliability of k-out-of-m structures in the independent exponential case. *Jour. Amer. Statist. Assoc.*, **63**, 850-854.

Bemis, B. M., Bain, L. J., and Higgins, J. J. (1972): Estimation and hypothesis testing for the parameters of a bivariate exponential distribution, *Jour. Amer. Statist. Assoc.*, **67**, 927-929.

Bhattacharyya, G. K., and Johnson, R. A. (1971): Maximum likelihood estimation and hypothesis testing in the bivariate exponential model of Marshall and Olkin. *Technical Report* No. 276, Department of Statistics, University of Wisconsin.

Block, H. W., and Basu, A. P. (1974): A continuous bivariate exponential extension. *Jour. Amer. Statist. Assoc.*, **69**, 1031-1037.

Freund, J. E. (1961): A bivariate extension of the exponential distribution. *Jour. Amer. Statist. Assoc.*, **56**, 971-977.

Friday, D. S., and Patil, G. P. (1977): A bivariate exponential model with applications to reliability and computer generation of random variables, in *The Theory and Application of Reliability*, eds., C. P. Tsokos and I. N. Shimi, New York, Academic Press, Inc., 527-549.

Gross, A. J. (1973): A competing risk model: a one organ subsystem plus a two organ subsystem. *IEEE Transactions on Reliability*, R-22, 24-27.

Gross, A. J., Clark, V. A., and Liu, V. (1971): Estimation of survival parameters when one of two organs must function for survival, *Biometrics*, **27**, 369-377.

Gross, A. J. and Lam, C. F. (1981): Paired observations from a survival distribution, *Biometrics*, **37**, 505-512.

Marshall, A. W., and Olkin, I. (1967): A multivariate exponential distribution. *Jour. Amer. Statist. Assoc.*, **62**, 30-40.

Mehrotra, K. G. and Michalek, J. E. (1976): Estimation of parameters and tests of independence in a continuous bivariate exponential distribution. unpublished.

Miller, R. (1977): The Jackknife–A Review; *Biometrika*, **61**, 1-15.

Proschan, F. and Sullo, P. (1976): Estimating the parameters of a multivariate exponential distribution. *Jour. Amer. Statist. Assoc.*, **71**, 465-472.

Pugh, E. L. (1963): The best estimate of reliability in the exponential case. *Jour. Oper. Res.*, **11**, 56-61.

Paper received : July, 1983.

Revised : July, 1984.

Printed by Haradhan Chakrabarti at Eka Press, 204/1, Barrackpore Trunk Road, Calcutta-35 and published by K. B. Goswami from Statistical Publishing Society, 204/1, Barrackpore Trunk Road, Calcutta 700 035.

APPENDIX M

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY NA	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE NA	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION The Ohio State University	6b. OFFICE SYMBOL *(If applicable)*	7a. NAME OF MONITORING ORGANIZATION AFOSR/NM
6c. ADDRESS (City, State, and ZIP Code) 1314 Kinnear Road Columbus, Ohio 43212		7b. ADDRESS (City, State, and ZIP Code) AFOSR/NM Bldg 410 Bolling AFB DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR	8b. OFFICE SYMBOL *(If applicable)* NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-82-0307

8c. ADDRESS (City, State, and ZIP Code)
AFOSR/NM
Bldg 410
Bolling AFB DC 20332-6448

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
61102F	2304		

11. TITLE *(Include Security Classification)* The Robustness of Several Estimators of the Survivorship Function with Randomly Censored Data

12. PERSONAL AUTHOR(S)
John P. Klein and Melvin L. Moeschberger

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 9-1-82 TO 12-31-87	May 31, 1988	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Censoring, parametric, nonparametric, product-limit
XXXX	XXXXXXXXXX	XXX	

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

The problem of estimating the survivorship function, $R(t)=P(T>t)$, arises frequently in both the engineering and biomedical sciences. In many applications the data one sees are censored due to the occurrence of some competing cause of failure such as withdrawal from the study, failure from some cause not under study, etc. In the biomedical sciences the distribution free estimator suggested by Kaplan and Meier (JASA 1958) is routinely used, while in the engineering sciences a parametric approach is more commonly used. In this report we study the efficiency of these two techniques when a particular parametric model such as the exponential, Weibull, normal, log normal, exponential power, Pareto, Gompertz, gamma, or bathtub shaped hazard distribution is assumed under a variety of censoring schemes and underlying failure models. We conclude that in most cases the parametric estimators outperform the distribution free estimator. The results are particularly striking if the Weibull forms of these estimators are used routinely.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL Maj. Brian Woodruff	22b. TELEPHONE (Include Area Code) (202) 767-5027	22c. OFFICE SYMBOL NM

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

THE ROBUSTNESS OF SEVERAL ESTIMATORS OF THE SURVIVORSHIP FUNCTION WITH RANDOMLY CENSORED DATA

by

John P. Klein
Department of Statistics
The Ohio State University

and

Melvin L. Moeschberger
Department of Preventive Medicine
The Ohio State University

Key Words: Censoring, Parametric, Nonparametric, Product-Limit

Abstract: The problem of estimating the survivorship function, $R(t)=P(T>t)$, arises frequently in both the engineering and biomedical sciences. In many applications the data one sees are censored due to the occurrence of some competing cause of failure such as withdrawal from the study, failure from some cause not under study, etc. In the biomedical sciences the distribution free estimator suggested by Kaplan and Meier (JASA 1958) is routinely used, while in the engineering sciences a parametric approach is more commonly used. In this report we study the efficiency of these two techniques when a particular parametric model such as the exponential, Weibull, normal, log normal, exponential power, Pareto, Gompertz, gamma, or bathtub shaped hazard distribution is assumed under a variety of censoring schemes and underlying failure models. We conclude that in most cases the parametric estimators outperform the distribution free estimator. The results are particularly striking if the Weibull forms of these estimators are used routinely.

I. Introduction

A common problem faced by the reliability analyst, as well as by the biomedical researcher, is to estimate the reliability or survivorship function $R(x) = P(X \geq x)$ for the time, X of occurrence of some event. This event may be time to failure of some system or time to failure of some component or subsystem of a larger more complex system in engineering applications. In biomedical applications, X may denote the time to death, relapse or cure of a patient, or time to death from a given cause or disease. Frequently, the observation of the event of interest is precluded by the occurrence of a censoring event at a random time Y. This censoring may be due to random observation periods or random entry into the study for each individual test unit. Random censoring may also be due to the failure of a system or individual due to an independent cause not under study. This censoring may often be quite heavy and the sample size on which inference is to be based quite small, particularly in the early exploratory period of product design or testing of a new therapeutic procedure.

Mathematically, the above problem is formulated as follows. Let X and Y be independent positive random variables with reliability functions $R(x) = P(X \geq x)$ and $G(y) = P(y \geq y)$, respectively. We do not observe X and Y directly but instead we observe T = minimum (X,Y) and

$$I = \begin{cases} 1 \text{ if } X < Y \text{ a failure} \\ 0 \text{ if } X \geq Y \text{ a censored observation} \end{cases}$$

Based on a random sample of size n we observe (T_j, I_j), $j=1, \ldots, n$. Our goal is to estimate $R(x)$.

There are two standard methodological approaches to estimating $R(x)$ in this framework. The first, which is used most commonly in biological and medical applications, is the nonparametric approach of Kaplan and Meier (1958). This estimator is constructed as follows. Suppose that the T_j's are ordered so that $T_1 < T_2 < \dots < T_n$ and the death times are unique. The estimator is defined by

$$(1.1) \qquad \hat{R}_{km}(x) = \begin{cases} 1 \text{ if } x \le T_1 \\ \prod_{k=1}^{i} \left(\frac{n-k}{n-k+1}\right)^{I_i} \text{ if } T_{i-1} < x \le T_i,\ i=1, \dots, n. \end{cases}$$

A special note is needed to cover the case where the largest observation is censored. Here $\hat{R}_{km}(x)$ is undefined for $t > T_n$. Efron (1967) suggests defining $\hat{R}_{km}(x) = 0$ for $t > T_n$, however we follow Gill (1980) who suggests estimating $R(x)$ by $\hat{R}_{km}(x) = \hat{R}(t_n)$ in this case (see Guerts (1987)). Under very general conditions on $R(x)$, $G(y)$ and an assumption of independence of X and Y, this estimator is consistent (c.f. Peterson (1977) and Winter (1987)); a generalized maximum likelihood estimator (c.f. Johansen (1978)); and converges weakly to a Gaussian process (c.f. Aalen (1976)). However recent studies by Geurts (1985) for nonproportional hazards censoring and Wellner (1985) (cf. Chen, Hollander and Langberg (1982)) for proportional hazards censoring have shown that for small sample sizes $\hat{R}_{km}$ is biased and that this bias is quite large for heavy censoring at median to small values of $R(t)$.

A second approach to estimating $R(x)$ is the fully parametric maximum likelihood approach which seems to be favored by researchers in the engineering sciences. Here a plausible model, $R_o(x;\underline{\theta})$, where $\underline{\theta} = (\theta_1, \dots, \theta_p)$ is a vector of unknown parameters, is postulated for $R(x)$. This model may, in some cases, be selected by some graphical technique (c.f. Nelson (1982)) or based on some theoretical grounds. Estimates of $\underline{\theta}$ are obtained by finding $\hat{\underline{\theta}}$ which maximizes the likelihood function

(1.2) $L(\underline{\theta}) = \prod_{i=1}^{n} h_o(T_i;\underline{\theta})^{I_i} R_o(T_i;\underline{\theta})$ where

$h(x) = -d\{\ln R(x)\}/dt$ is the hazard function of X. The resulting estimator of R(x) is $\hat{R}_{MLE}(x) = R_o(x;\hat{\underline{\theta}})$ which, under very general regularity conditions, is asymptotically unbiased, consistent, efficient, and normally distributed (c.f. Bain (1978)) if the model R_o is properly selected. However its small sample properties and the robustness of this estimator to misspecification of the model may be suspect.

The goal of this paper is to explore the small and moderate sample size performance of the nonparametric and parametric estimators of R(x) under light to heavy censoring by means of a Monte Carlo study. For R_o we study the exponential, Weibull, normal, lognormal, exponential power, log logistic, Pareto and Gompertz models. Data is simulated from a variety of distributional shapes including those with constant, increasing, decreasing, or bathtub or U shaped hazard functions.

II. Parametric Models for Reliability

In this section we describe the models used for $R_o(\cdot|\underline{\theta})$ in our Monte Carlo study. The first is the exponential distribution with $R_o(x|\theta) = \exp(-x/\theta)$. This model, which has a constant hazard rate $h(x) = 1/\theta$ has been extensively used and studied in this context. For example, Davis (1952) used this model to study the lifetimes of manufactured items while Feigel and Zelen (1965)) suggest its use in modeling survival or remission times for a chronic illness.

The second model we have considered is the two parameter Weibull distribution with reliability function $R(x|\alpha,\beta) = \exp(-(t/\alpha)^{\beta}), \alpha,\beta>0$. This flexible model has been used in reliability (c.f. Weibull (1951)), medical (c.f. Whittemore and Altschuler (1976)), and animal (Pike (1966)) studies. It has great flexibility in that its hazard rate can be monotone decreasing ($\beta < 1$), constant ($\beta = 1$), or monotone increasing ($\beta > 1$).

The next model considered is the normal distribution with mean μ and variance σ^2. This model, which has an increasing hazard rate, is included due to its naive use by many not versed in the reliability literature. We can note, however, that Davis (1952) and Barlow and Proschan (1965) have suggested it as a model for lifetime data. We also consider the related log normal distribution with reliability function $R_o(x|a,s) = 1-\Phi((\ln x-a)/s))$, $s > o$, $-\infty < a < \infty$. This distribution has a humpshaped hazard rate. Its use in life studies has been suggested by a number of researchers including Nelson and Hahn (1972) in an engineering context and Whittemore and Altschuler (1976) in a medical context. Estimates of μ and σ or a and s were obtained by the EM algorithm (see Lawless (1982)).

The next distribution considered was the exponential power distribution proposed by Smith and Bain (1975). This distribution has reliability function.

$$R(x|\alpha,\beta) = \exp\left(1-e^{(x/\alpha)^{\beta}}\right),\ x,\ \alpha,\ \beta > o$$

and hazard rate $h(x) = \beta\, x^{\beta-1}\exp((x/\alpha)^{\beta})/\alpha^{\beta}$.

It was chosen due to its flexibility since it allows for U shaped hazard rates when $\beta < 1$ and monotone increasing hazard rates when $\beta > 1$.

The sixth distribution studied is the log logistic distribution with reliability function $R(x|\alpha,\beta) = 1/(1+\alpha x^{\beta})$, α, $\beta > o$. This model, with hazard rate $\alpha\beta x^{\beta-1}/(1+\alpha x^{\beta})$, behaves like the log normal with a humpshaped hazard rate for $\beta > 1$ and has a monotone decreasing hazard rate for $\beta < 1$.

The seventh distribution considered is the Pareto, with reliability function $R(x|\alpha,\beta) = (1+\alpha x)^{-\beta}$, $x, \alpha, \beta > o$, and hazard rate $h(x) = \beta\alpha/(1+\alpha x)$ which is strictly decreasing. This model arises in modeling a heterogeneous exponential population as follows . Suppose that X has an exponential distribution with random hazard rate λ. If λ follows a gamma distribution with density $g(\lambda) = \lambda^{\beta-1} \exp(-\lambda/\alpha)/(\alpha^{\beta}\Gamma(\beta))$ then (unconditionally) X has a Pareto distribution. Maximum likelihood estimation of α and β are on the boundary of the parameter space whenever

$$(\sum_{j=1}^{n} I_j) \sum_{j=1}^{n} (T_j I_j)^2 - 2 (\sum_{j=1}^{n} (T_j I_j))^2 < 0,$$

in which case the estimated reliability reduces to that of the exponential (c.f. Lee and Klein (1988)).

The final distribution considered is the Gompertz distribution with reliability function $R_0(x|\alpha,\beta) = \exp(\alpha(1-e^{\beta x})/\beta)$, $\alpha, \beta, x > o$ which has an exponentially increasing hazard rate $h(x) = \alpha e^{\beta x}$. This model has been used extensively in modeling mortality data (see Elandt-Johnson and Johnson (1980) and Gehan and Siddiqui (1973)).

3. Simulation Study

To study the performance of the maximum likelihood estimators of R(x) for the above models and of the Kaplan-Meier Product Limit Estimator 1000 samples of size 25 and 50, with 0%, 10%, 30%, or 50% of each sample being randomly censored, were generated from the following populations of failure times: (In each case we fixed the mean life at 1 to make comparisons easier).

1) exponential;

2) Weibull with β = .5, 2, 4, 8;

3) normal with σ = .05, .1, .15;

4) log normal with s = .37, .51, .61;

5) exponential power with β = .25, .50, 1, 8;

7) Pareto with β = 1, 2, 4;

8) Gompertz with β = .5, 1, 2;

9) gamma distribution with shape parameter β = .5, 2, 4, 8;

10) A bathtub shape hazard distribution proposed by Glaser (1980), which is a mixture of a gamma with shape parameter 3 with probability p and an exponential (with the same scale parameter) with probability q = (1-p), p = .15, .25, .4, and .6; and

11) log logistic with β = 2, 3, 4.

Censoring random variables were generated from an exponential distribution with the appropriate parameter for all death distributions. Additionally, proportional hazards censoring was used for the Weibull and exponential power distribution. Exponential censoring will give a censoring pattern with heavy early censoring for distributions with an increasing hazard rate and heavy late censoring for distributions with a decreasing failure rate. For bathtub shaped hazard rates the censoring will increase to a maximum and then become light for large observed times.

As a measure of the performance of the eight estimators of R(t) we consider an estimator of the integrated mean squared error defined as

(3.1)

$$\text{IMSE}\ (\hat{R}) = E\ \{\int_0^\infty (\hat{R}(t) - R(t))^2\, dt\}$$

where R(t) is the true reliability function. We estimate this quantity by

(3.2)

$$\text{EIMSE}\ (\hat{R}) = \sum_{j=1}^{1000} \int_0^{t_{.95}} (\hat{R}_j(t) - R(t))^2\, dt/1000$$

where $P(T \leq t_p) = p$ and $\hat{R}_j(x)$ is the estimator of survival on the j^{th} simulation. The results, reported in Tables 1-8, are the values of the ratio of integrated mean square error when the product limit estimator is used to the given maximum likelihood method (i.e., EIMSE(KM)/EIMSE(MLE)), so that a value greater than one implies that the corresponding likelihood method performs better than the Kaplan-Meier estimator. In Figures 1-7 we present plots of the relative mean squared error (MSE(KM)/MSE(MLE)) at the 5th to 95th percentiles of the true distribution for a set of representative distributions based on a sample of size 25 with 30 percent censoring. These distributions are the exponential, with constant hazard rate, (Figure 1); the gamma with shape parameter .5 (Figure 2), with a decreasing hazard rate; the Weibull with β = 8 (Figure 3) and the Gompertz with β = 2 (Figure 4), both with increasing hazard rate; the log-logistic with β = 3 (Figure 5), which has a humpshaped hazard rate; and Glaser's bathtub shape hazard distribution with p = .6 (Figure 6) and the exponential power distribution with β = .25 (Figure 7), both with U shape hazard rates. Exponential censoring was used throughout.

4. Discussion

Miller (1983) noted that, asymptotically, the efficiency of the Kaplan-Meier estimator is quite low compared to the parametric maximum likelihood estimator under the assumption the parametric estimation model is correctly chosen. He showed that this was particularly true for high censoring proportions and reliabilities estimated in either tail. Of course when the parametric model is chosen incorrectly the maximum likelihood estimator is asymptotically inefficient. Our results show that for small to moderate sample sizes the parametric estimators outperform the Kaplan-Meier estimator not only when the parametric model is chosen correctly but also for families of models with similar shapes. The parametric models tend to do even better as the percentage of censored observations increases, reflecting the higher bias of the product limit estimator.

Specific recommendations can be made based on these tables and graphs for the use of certain parametric models. First, notice the poor performance of the Pareto maximum likelihood estimator. Its integrated mean squared error is worse than that of the Kaplan-Meier estimate in 286 out of 304 cases. This may be due, in part, to the instability of its parameter estimates. The exponential MLE also does poorly except for Glaser's bathtub shape distribution and the exponential. Its use as a routine model for reliability is not indicated by these results. For the remaining models considered, the results are mixed. For distributions with a decreasing hazard the Weibull and log logistic MLE's seem to perform well. The exponential power distribution MLE performs well for the Weibull and gamma models with $\beta < 1$. The use of the Gompertz, normal and log normal MLE's is clearly not indicated. For distributions with a bathtub shaped hazard rate (Glaser's distribution, exponential power distribution with $\beta < 1$) the Weibull and the exponential power distributions outperform the Kaplan-Meier estimator, while the remaining distributions do not. For humpshaped hazard rate distributions (log normal, log logistic) the use of the log logistic or log normal MLE is indicated. One should note the relatively poor performance of the Weibull MLE here. The use of the log logistic is, perhaps, indicated due to its simpler computation form.

For distributions with an increasing failure rate the picture is not so clear cut. The Weibull MLE outperforms the product limit estimator except when the true model is normal or Gompertz with an extremely steep hazard rate ($\beta = 2$). The exponential power distribution does well except for normal data. The Gompertz does well except for normal data and for Weibull data with a relatively flat hazard rate ($\beta = 2$). The log normal and log logistic MLE's provide a reasonable estimator of the reliability when the data is Weibull or normal. The Gompertz MLE is good for distributions with a steep hazard rate such as the Gompertz, exponential power distribution and Weibull with shape parameter 8. The estimator which most consistently outperforms the Kaplan-Meier estimator for increasing hazard rate distributions is the Weibull which wins in 131 of 142 cases considered.

The above discussion suggests that a statistician armed with the Weibull, log logistic and exponential power distribution MLE's can provide better estimates of the reliability function than one armed only with the Kaplan-Meier estimator. By a preliminary graphical look at the hazard rate (c.f. Nelson (1982)) he or she can get a crude idea of shape of the hazard rate and pick the most appropriate model of these three.

Acknowledgments

The authors wish to thank Mr. Paul Hshieh and Ms. Jane Chang for their programming assistance. This research was supported by the Air Force Office of Scientific Research under contract AFOSR-82-0307.

References

1. Aalen, O. (1976). "Nonparametric inference in connection with multiple decrement models." Scandinavian Journal of Statistics 3, 15-17.

2. Bain, L.J. (1978). Statistical Analysis of Reliability and Life Test Models, Marcel-Dekker, New York.

3. Barlow, R.E., and Proschan, F. (1965). Mathematical Theory of Reliability, Wiley, New York.

4. Chen, Y.Y., Hollander, M., and Langberg, N.A. (1982). "Small sample results for the Kaplan-Meier estimator." Journal of the American Statistical Association 67, 141-144.

5. Davis, D.J. (1952). "An analysis of some failure data." Journal of the American Statistical Association, 47, 113-150.

6. Efron, B. (1967). "The two sample problem with censored data." Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics IV, Prentice-Hall, New York, 831-853.

7. Elandt-Johnson, R.C. and Johnson, N.L. (1980). Survival Models and Data Analyses, Wiley, New York.

8. Feigl, P. and Zelen, M. (1965). "Estimation of exponential survival probabilities with concomitant information." Biometrics 21, 826-838.

9. Gehan, R.E. and Siddiqui, M.M. (1973). "Simple regression methods for survival time studies." Journal of the American Statistical Association 68, 848-856.

10. Geurts, J.H.J. (1985). "Some small sample non-proportional hazards results for the Kaplan-Meier Estimator." Statistical Neerlandica 39, 1-13.

11. Geurts, J.H.J. (1987). "On the small-sample performance of Efron's and of Gill's version of the product limit estimator under nonproportional hazards. Biometrics 43, 683-692.

12. Gill, R.D. (1980). Censoring and Stochastic Integrals, Mathematical Center, Amsterdam.

13. Glaser, R. (1980). "Bathtub and related failure rate characteristics." Journal of the American Statistical Association 75, 667-672.

14. Johansen, S. (1978). "The product limit estimator as maximum likelihood estimator." Scandinavian Journal of Statistics 5, 195-199.78).

15. Kaplan. E.L. and Meier, P. (1958). "Nonparametric estimation from incomplete observations" Journal of the American Statistical Association 53, 457-481.

16. Lawless, J.F. (1982). Statistical Models and Methods for Lifetime Data, Wiley, New York.

17. Lee, S. and Klein, J.P. (1988). "Statistical methods for combining laboratory and field data based on a random environmental stress model." Proceedings of the OSU-Korea Conference on Recent Developments in Statistics and their Applications (To Appear).

18. Miller, R.G. (1983). "What price Kaplan-Meier?" Biometrics 39, 1077-1081.

19. Nelson, W. (1982). Applied Life Data Analysis. Wiley, New York.

20. Nelson, W. and Hahn, G.I. (1972). "Linear estimation of a regression relationship with censored data-Part I. Simple Methods and their Applications." Technometrics 14, 247-269.

21. Peterson, A.V. (1977). "Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions." Journal of the American Statistical Association 72, 854-858.

22. Pike, M.C. (1966). "A method of analysis of a certain class of experiments in carcinogenesis." Biometrics 22, 142-161.

23. Smith, R.M. and Bain, L.J. (1975). "An exponential power life-testing distribution." Communications in Statistics-Theory and Methods A4(5), 469-481.

24. Weibull, W. (1951). "A statistical distribution function of wide applicability." Journal Appl. Mech. 18, 293-297.

25. Wellner, J. (1985). "A heavy censoring limit theorem for the product limit estimator." Annals of Statistics 13:150-162.

26. Whittemore, A and Altschuler, B. (1976). "Lung cancer incidence in cigarette smokers: further analysis of Doll and Hill's data for British physicians." Biometrics 32, 805-816.

27. Winter, B.B. (1987). "Nonparametric estimation with censored data from a distribution with nonincreasing density." Communications in Statistics-Theory and Methods 16(1), 93-120.

TABLE 1 - EXPONENTIAL MLE

DISTRIBUTION OF DEATHS	CENSORING DISTRIBUTION	SAMPLE SIZE 25 PERCENT CENSORED 0%	10%	30%	50%	SAMPLE SIZE 50 PERCENT CENSORED 0%	10%	30%	50%
EXPONENTIAL	EXP	1.96	1.91	2.11	2.85	1.99	1.92	2.05	2.84
WEIBULL,B=0.50	EXP	0.26	0.44	1.10	3.17	0.03	0.41	0.62	2.03
WEIBULL,B=2.00	EXP	0.29	0.34	0.47	0.61	0.16	0.18	0.23	0.32
WEIBULL,B=4.00	EXP	0.06	0.07	0.10	0.16	0.16	0.03	0.05	0.08
WEIBULL,B=8.00	EXP	0.02	0.03	0.04	0.06	0.01	0.01	0.02	0.03
WEIBULL,B=0.50	WEI		0.35	0.66	1.44		0.21	0.40	0.85
WEIBULL,B=2.00	WEI		0.34	0.45	0.58		0.18	0.25	0.34
WEIBULL,B=4.00	WEI		0.07	0.11	0.19		0.03	0.05	0.10
WEIBULL,B=8.00	WEI		0.03	0.04	0.08		0.01	0.02	0.04
NORMAL,σ=0.05	EXP	0.01	0.01	0.02	0.03	0.00	0.01	0.01	0.02
NORMAL,σ=0.10	EXP	0.02	0.02	0.03	0.05	0.01	0.01	0.01	0.02
NORMAL,σ=0.15	EXP	0.02	0.03	0.04	0.07	0.01	0.01	0.02	0.03
LN NORMAL,S=.37	EXP	0.09	0.10	0.15	0.21	0.04	0.05	0.07	0.10
LN NORMAL,S=.51	EXP	0.20	0.23	0.29	0.38	0.10	0.12	0.15	0.21
LN NORMAL,S=.61	EXP	0.34	0.38	0.48	0.58	0.18	0.20	0.26	0.32
EP,B=0.50	EXP	0.93	0.98	1.29	2.54	0.63	0.67	0.89	1.94
EP,B=1.00	EXP	1.08	1.31	1.49	1.67	0.66	0.76	1.04	1.17
EP,B=8.00	EXP	0.02	0.02	0.03	0.06	0.01	0.01	0.02	0.03
EP,B=0.25	EP		0.24	0.50	1.26		0.14	0.28	0.67
EP,B=0.50	EP		0.97	1.22	1.82		0.69	0.84	1.30
EP,B=1.00	EP		1.21	1.42	1.54		0.78	1.00	1.23
EP,B=8.00	EP		0.02	0.03	0.07		0.01	0.02	0.03
EP,B=0.25	EXP	0.19	0.31	1.04	3.32	0.11	0.17	0.62	2.50
PARETO,B=1.00	EXP	0.34	0.70	1.56	3.86	0.22	0.52	1.11	2.76
PARETO,B=2.00	EXP	0.72	1.02	1.73	3.48	0.59	0.80	1.34	2.77
PARETO,B=4.00	EXP	1.12	1.43	1.87	3.23	0.99	1.20	1.58	2.98
GOMPERTZ,B=0.50	EXP	1.31	1.40	1.62	1.75	0.82	0.97	1.17	1.39
GOMPERTZ,B=1.00	EXP	0.61	0.71	0.92	1.14	0.32	0.38	0.53	0.70
GOMPERTZ,B=2.00	EXP	0.22	0.28	0.37	0.55	0.11	0.14	0.20	0.29
GAMMA,B=0.50	EXP	0.66	0.81	1.22	2.74	0.44	0.52	0.82	1.94
GAMMA,B=2.00	EXP	2.59	0.87	1.00	1.12	2.46	0.50	0.61	0.69
GAMMA,B=4.00	EXP	0.53	0.24	0.32	0.43	0.49	0.12	0.17	0.23
GAMMA,B=8.00	EXP	1.15	0.71	0.14	0.20	1.15	0.05	0.07	0.10
BATHTUB,P=0.15	EXP	1.83	1.88	1.93	2.95	1.87	1.76	1.89	2.89
BATHTUB,P=0.25	EXP	1.95	1.88	2.00	2.64	1.90	1.90	2.01	2.91
BATHTUB,P=0.40	EXP	2.09	2.07	2.04	2.49	1.97	2.00	1.99	2.47
BATHTUB,P=0.60	EXP	1.64	1.71	1.82	1.96	1.18	1.29	1.47	1.68
LN LOGISTIC,B= 2.	EXP	0.62	0.80	1.18	1.56	0.37	0.59	0.86	1.21
LN LOGISTIC,B= 3.	EXP	0.27	0.30	0.39	0.46	0.15	0.16	0.21	0.27
LN LOGISTIC,B= 4.	EXP	0.12	0.14	0.19	0.25	0.06	0.07	0.09	0.14

TABLE 2 - WEIBULL MLE

DISTRIBUTION OF DEATHS	CENSORING DISTRIBUTION	SAMPLE SIZE 25 PERCENT CENSORED 0%	10%	30%	50%	SAMPLE SIZE 50 PERCENT CENSORED 0%	10%	30%	50%
EXPONENTIAL	EXP	1.49	1.51	1.67	2.13	1.51	1.52	1.66	2.15
WEIBULL,B=0.50	EXP	1.49	1.55	2.15	4.60	1.51	1.53	2.38	5.76
WEIBULL,B=2.00	EXP	1.51	1.52	1.56	1.67	1.55	1.53	1.61	1.69
WEIBULL,B=4.00	EXP	1.61	1.61	1.54	1.62	1.65	1.64	1.63	1.60
WEIBULL,B=8.00	EXP	1.67	1.68	1.71	1.67	1.80	1.77	1.70	1.72
WEIBULL,B=0.50	WEI		1.57	1.70	2.64		1.59	1.72	2.50
WEIBULL,B=2.00	WEI		1.53	1.63	1.91		1.52	1.59	1.96
WEIBULL,B=4.00	WEI		1.60	1.66	1.84		1.62	1.63	1.84
WEIBULL,B=8.00	WEI		1.71	1.65	1.82		1.67	1.73	1.77
NORMAL,σ=0.05	EXP	0.82	0.95	1.14	1.31	0.49	0.54	0.66	0.90
NORMAL,σ=0.10	EXP	1.12	1.16	1.37	1.56	0.79	0.85	0.99	1.22
NORMAL,σ=0.15	EXP	1.30	1.40	1.46	1.56	1.08	1.13	1.25	1.40
LN NORMAL,S=.37	EXP	0.98	1.01	1.13	1.26	0.71	0.77	0.86	1.04
LN NORMAL,S=.51	EXP	0.97	1.06	1.15	1.30	0.75	0.81	0.95	1.15
LN NORMAL,S=.61	EXP	1.02	1.01	1.16	1.39	0.77	0.83	0.99	1.23
EP,B=0.25	EXP	1.75	1.60	2.19	4.11	1.67	1.50	2.18	5.36
EP,B=0.50	EXP	1.50	1.57	1.69	2.33	1.43	1.43	1.58	2.34
EP,B=1.00	EXP	1.38	1.47	1.62	1.83	1.29	1.33	1.57	1.67
EP,B=8.00	EXP	1.60	1.50	1.68	1.60	1.34	1.41	1.51	1.52
EP,B=0.25	EP		1.75	1.71	2.12		1.61	1.66	1.80
EP,B=0.50	EP		1.52	1.70	1.93		1.52	1.53	1.80
EP,B=1.00	EP		1.42	1.59	1.73		1.40	1.46	1.74
EP,B=8.00	EP		1.49	1.62	1.69		1.41	1.52	1.55
PARETO,B=1.00	EXP	1.04	1.21	1.66	3.14	0.88	1.11	1.50	2.95
PARETO,B=2.00	EXP	1.20	1.31	1.63	2.60	1.20	1.21	1.47	2.47
PARETO,B=4.00	EXP	1.34	1.43	1.63	2.42	1.31	1.34	1.51	2.26
GOMPERTZ,B=0.50	EXP	1.42	1.45	1.59	1.82	1.33	1.43	1.53	1.70
GOMPERTZ,B=1.00	EXP	1.30	1.30	1.44	1.62	1.07	1.12	1.26	1.56
GOMPERTZ,B=2.00	EXP	1.09	1.14	1.25	1.42	0.91	0.95	1.04	1.25
GAMMA,B=0.50	EXP	1.54	1.51	1.73	2.85	1.52	1.55	1.69	2.55
GAMMA,B=2.00	EXP	1.42	1.43	1.56	1.68	1.41	1.39	1.47	1.66
GAMMA,B=4.00	EXP	1.38	1.41	1.43	1.56	1.26	1.27	1.34	1.43
GAMMA,B=8.00	EXP	1.28	1.30	1.37	1.42	1.06	1.08	1.23	1.34
BATHTUB,P=0.15	EXP	1.47	1.52	1.62	2.29	1.60	1.51	1.63	2.23
BATHTUB,P=0.25	EXP	1.50	1.51	1.65	2.10	1.51	1.56	1.71	2.30
BATHTUB,P=0.40	EXP	1.50	1.53	1.67	2.02	1.47	1.54	1.66	2.03
BATHTUB,P=0.60	EXP	1.43	1.49	1.59	1.77	1.38	1.41	1.51	1.70
LN LOGISTIC,B= 2.	EXP	0.77	0.87	1.18	1.74	0.51	0.67	0.98	1.47
LN LOGISTIC,B= 3.	EXP	0.77	0.84	1.03	1.28	0.52	0.60	0.81	1.08
LN LOGISTIC,B= 4.	EXP	0.78	0.84	0.98	1.15	0.48	0.56	0.71	0.97

TABLE 3 - LOG NORMAL MLE

DISTRIBUTION OF DEATHS	CENSORING DISTRIBUTION	SAMPLE SIZE 25 PERCENT CENSORED 0%	10%	30%	50%	SAMPLE SIZE 50 PERCENT CENSORED 0%	10%	30%	50%
EXPONENTIAL	EXP	1.30	1.31	1.37	1.49	0.93	0.98	0.93	1.04
WEIBULL,B=0.50	EXP	1.43	1.25	1.28	2.33	1.05	1.21	0.87	1.80
WEIBULL,B=2.00	EXP	1.13	1.17	1.30	1.43	0.91	0.92	1.02	1.08
WEIBULL,B=4.00	EXP	1.10	1.15	1.22	1.32	0.86	0.92	0.94	1.12
WEIBULL,B=8.00	EXP	1.04	1.10	1.17	1.23	0.87	0.88	0.96	1.05
WEIBULL,B=0.50	WEI		1.39	1.15	1.53		0.94	0.83	0.92
WEIBULL,B=2.00	WEI		1.17	1.25	1.44		0.94	0.97	1.05
WEIBULL,B=4.00	WEI		1.14	1.24	1.42		0.91	1.01	1.04
WEIBULL,B=8.00	WEI		1.15	1.20	1.39		0.90	1.02	1.06
NORMAL,σ=0.05	EXP	1.79	1.77	1.80	1.80	1.72	1.77	1.70	1.82
NORMAL,σ=0.10	EXP	1.58	1.65	1.73	1.65	1.52	1.55	1.53	1.60
NORMAL,σ=0.15	EXP	1.47	1.47	1.55	1.56	1.36	1.37	1.41	1.53
LN NORMAL,S=.37	EXP	1.56	1.53	1.64	1.64	1.56	1.58	1.58	1.68
LN NORMAL,S=.51	EXP	1.50	1.54	1.55	1.73	1.59	1.32	1.61	1.75
LN NORMAL,S=.61	EXP	1.54	1.51	1.62	1.82	1.56	1.16	1.64	1.83
EP,B=0.25	EXP	1.06	0.78	0.94	1.88	0.62	0.45	0.57	1.55
EP,B=0.50	EXP	1.04	1.01	0.91	1.18	0.67	0.61	0.55	0.76
EP,B=1.00	EXP	0.85	0.99	1.08	1.22	0.57	0.62	0.70	0.69
EP,B=8.00	EXP	0.84	0.84	1.00	1.05	0.59	0.61	0.75	0.81
EP,B=0.25	EP		0.92	0.75	0.90		0.54	0.41	0.47
EP,B=0.50	EP		1.03	0.97	1.05		0.65	0.57	0.61
EP,B=1.00	EP		0.91	1.02	1.07		0.61	0.65	0.71
EP,B=8.00	EP		0.85	0.98	1.11		0.63	0.73	0.76
PARETO,B=1.00	EXP	1.53	1.45	1.62	2.53	1.38	1.34	1.41	1.99
PARETO,B=2.00	EXP	1.44	1.48	1.54	1.97	1.28	1.21	1.16	1.45
PARETO,B=4.00	EXP	1.35	1.43	1.40	1.76	1.18	1.12	1.06	1.23
GOMPERTZ,B=0.50	EXP	0.96	0.99	1.10	1.17	0.62	0.67	0.70	0.74
GOMPERTZ,B=1.00	EXP	0.75	0.78	0.91	1.05	0.45	0.47	0.55	0.62
GOMPERTZ,B=2.00	EXP	0.56	0.63	0.70	0.86	0.33	0.37	0.42	0.50
GAMMA,B=0.50	EXP	1.11	1.02	0.94	1.38	0.72	0.66	0.55	0.84
GAMMA,B=2.00	EXP	1.37	1.41	1.56	1.69	1.18	1.20	1.32	1.31
GAMMA,B=4.00	EXP	1.43	1.53	1.58	1.67	1.38	1.34	1.49	1.58
GAMMA,B=8.00	EXP	1.53	0.78	1.58	1.66	1.45	1.48	1.54	1.64
BATHTUB,P=0.15	EXP	1.35	1.33	1.33	1.62	1.02	1.04	1.00	1.09
BATHTUB,P=0.25	EXP	1.29	1.31	1.37	1.56	0.96	1.00	1.01	1.10
BATHTUB,P=0.40	EXP	1.12	1.21	1.27	1.45	0.79	0.87	0.90	0.95
BATHTUB,P=0.60	EXP	0.97	1.00	1.08	1.13	0.61	0.64	0.67	0.69
LN LOGISTIC,B= 2.	EXP	1.34	1.38	1.66	2.02	1.22	1.36	1.50	2.02
LN LOGISTIC,B= 3.	EXP	1.31	1.36	1.45	1.67	1.28	1.28	1.39	1.58
LN LOGISTIC,B= 4.	EXP	1.34	1.33	1.44	1.49	1.26	1.32	1.35	1.53

TABLE 4 - NORMAL MLE

DISTRIBUTION OF DEATHS	CENSORING DISTRIBUTION	SAMPLE SIZE 25 PERCENT CENSORED				SAMPLE SIZE 50 PERCENT CENSORED			
		0%	10%	30%	50%	0%	10%	30%	50%
EXPONENTIAL	EXP	0.56	0.61	0.87	1.39	0.34	0.41	0.58	0.99
WEIBULL,B=0.50	EXP	0.16	0.28	0.87	2.54	0.08	0.26	0.50	1.68
WEIBULL,B=2.00	EXP	1.30	1.31	1.37	1.52	1.13	1.13	1.20	1.35
WEIBULL,B=4.00	EXP	1.62	1.63	1.60	1.68	1.65	1.65	1.65	1.65
WEIBULL,B=8.00	EXP	1.38	1.45	1.51	1.53	1.32	1.31	1.36	1.45
WEIBULL,B=0.50	WEI		0.21	0.38	1.01		0.11	0.21	0.60
WEIBULL,B=2.00	WEI		1.35	1.48	1.88		1.16	1.29	1.69
WEIBULL,B=4.00	WEI		1.63	1.73	2.00		1.61	1.70	1.99
WEIBULL,B=8.00	WEI		1.49	1.52	1.76		1.31	1.46	1.57
NORMAL,σ=0.05	EXP	1.83	1.81	1.83	1.82	1.79	1.79	1.74	1.84
NORMAL,σ=0.10	EXP	1.66	1.73	1.80	1.73	1.67	1.71	1.65	1.70
NORMAL,σ=0.15	EXP	1.63	1.64	1.69	1.69	1.63	1.63	1.64	1.72
LN NORMAL,S=.37	EXP	0.92	0.95	1.08	1.23	0.68	0.74	0.82	1.01
LN NORMAL,S=.51	EXP	0.66	0.76	0.90	1.09	0.45	0.51	0.65	0.87
LN NORMAL,S=.61	EXP	0.57	0.60	0.78	1.07	0.35	0.40	0.55	0.79
EP,B=0.25	EXP	0.14	0.24	1.03	3.17	0.07	0.14	0.64	2.44
EP,B=0.50	EXP	0.46	0.52	0.79	1.60	0.27	0.32	0.51	1.12
EP,B=1.00	EXP	1.19	1.15	1.25	1.43	0.90	0.92	1.02	1.16
EP,B=8.00	EXP	1.13	1.11	1.30	1.31	0.84	0.88	1.02	1.10
EP,B=0.25	EP		0.17	0.33	0.95		0.09	0.18	0.54
EP,B=0.50	EP		0.51	0.69	1.16		0.31	0.44	0.78
EP,B=1.00	EP		1.19	1.26	1.63		0.92	1.01	1.32
EP,B=8.00	EP		1.13	1.28	1.43		0.90	1.03	1.12
PARETO,B=1.00	EXP	0.17	0.28	0.68	1.82	0.08	0.16	0.39	1.06
PARETO,B=2.00	EXP	0.26	0.35	0.70	1.58	0.15	0.21	0.40	0.92
PARETO,B=4.00	EXP	0.34	0.45	0.72	1.51	0.21	0.27	0.45	0.90
GOMPERTZ,B=0.50	EXP	1.08	1.08	1.15	1.50	0.82	0.83	0.92	1.10
GOMPERTZ,B=1.00	EXP	1.39	1.41	1.43	1.52	1.22	1.26	1.32	1.39
GOMPERTZ,B=2.00	EXP	1.48	1.55	1.61	1.77	1.42	1.45	1.57	1.72
GAMMA,B=0.50	EXP	0.33	0.45	0.79	1.78	0.20	0.25	0.48	1.18
GAMMA,B=2.00	EXP	0.81	0.86	1.03	1.24	0.55	0.63	0.72	0.97
GAMMA,B=4.00	EXP	1.04	1.10	1.18	1.38	0.80	0.86	0.97	1.11
GAMMA,B=8.00	EXP	1.25	1.41	1.35	1.42	1.05	1.06	1.22	1.33
BATHTUB,P=0.15	EXP	0.52	0.61	0.82	1.44	0.33	0.38	0.55	0.92
BATHTUB,P=0.25	EXP	0.60	0.65	0.84	1.27	0.38	0.42	0.57	0.91
BATHTUB,P=0.40	EXP	0.72	0.78	0.93	1.36	0.46	0.55	0.67	0.96
BATHTUB,P=0.60	EXP	0.91	0.99	1.11	1.41	0.65	0.73	0.84	1.09
LN LOGISTIC,B= 2.	EXP	0.24	0.31	0.60	1.17	0.11	0.18	0.39	0.78
LN LOGISTIC,B= 3.	EXP	0.40	0.47	0.70	1.00	0.23	0.28	0.47	0.74
LN LOGISTIC,B= 4.	EXP	0.57	0.65	0.84	1.04	0.34	0.41	0.57	0.84

TABLE 5 - EXPONENTIAL POWER MLE

DISTRIBUTION OF DEATHS	CENSORING DISTRIBUTION	SAMPLE SIZE 25 PERCENT CENSORED				SAMPLE SIZE 50 PERCENT CENSORED			
		0%	10%	30%	50%	0%	10%	30%	50%
EXPONENTIAL	EXP	1.29	1.37	1.54	1.95	1.14	1.31	1.46	1.96
WEIBULL,B=0.50	EXP	1.18	1.41	2.11	4.77	1.07	1.37	2.01	4.88
WEIBULL,B=2.00	EXP	1.34	1.40	1.41	1.55	1.19	1.24	1.32	1.47
WEIBULL,B=4.00	EXP	1.41	1.42	1.42	1.51	1.15	1.18	1.28	1.33
WEIBULL,B=8.00	EXP	1.34	1.36	1.49	1.55	1.07	1.05	1.21	1.35
WEIBULL,B=0.50	WEI		1.34	1.56	2.44		1.31	1.43	2.05
WEIBULL,B=2.00	WEI		1.42	1.53	1.84		1.28	1.41	1.78
WEIBULL,B=4.00	WEI		1.44	1.55	1.80		1.24	1.39	1.69
WEIBULL,B=8.00	WEI		1.44	1.49	1.74		1.15	1.36	1.63
NORMAL,σ=0.05	EXP	0.36	0.35	0.49	0.70	0.17	0.13	0.20	0.32
NORMAL,σ=0.10	EXP	0.56	0.50	0.72	1.08	0.28	0.25	0.37	0.56
NORMAL,σ=0.15	EXP	0.70	0.71	0.93	1.20	0.42	0.42	0.56	0.77
LN NORMAL,S=.37	EXP	0.58	0.63	0.77	0.97	0.31	0.37	0.46	0.61
LN NORMAL,S=.51	EXP	0.60	0.72	0.84	1.01	0.34	0.42	0.55	0.76
LN NORMAL,S=.61	EXP	0.65	0.70	0.87	1.13	0.36	0.44	0.59	0.84
EP,B=0.25	EXP	1.53	1.55	2.79	6.73	1.55	1.53	3.11	9.26
EP,B=0.50	EXP	1.51	1.61	1.81	2.78	1.57	1.59	1.77	3.12
EP,B=1.00	EXP	1.57	1.59	1.69	1.80	1.58	1.59	1.73	1.84
EP,B=8.00	EXP	1.89	1.77	1.96	1.77	1.89	1.85	1.88	1.83
EP,B=0.25	EP		1.55	1.67	2.74		1.51	1.75	2.40
EP,B=0.50	EP		1.54	1.79	2.14		1.65	1.68	2.18
EP,B=1.00	EP		1.57	1.65	1.90		1.67	1.64	1.99
EP,B=8.00	EP		1.73	1.77	1.82		1.81	1.75	1.76
PARETO,B=1.00	EXP	0.64	0.92	1.41	2.82	0.41	0.70	1.12	2.18
PARETO,B=2.00	EXP	0.84	1.02	1.39	2.37	0.63	0.81	1.12	1.91
PARETO,B=4.00	EXP	1.00	1.18	1.42	2.27	0.77	0.97	1.22	1.82
GOMPERTZ,B=0.50	EXP	1.55	1.55	1.62	1.82	1.58	1.65	1.65	1.75
GOMPERTZ,B=1.00	EXP	1.50	1.49	1.56	1.71	1.43	1.46	1.51	1.78
GOMPERTZ,B=2.00	EXP	1.32	1.35	1.42	1.54	1.24	1.26	1.31	1.50
GAMMA,B=0.50	EXP	1.43	1.51	1.82	3.28	1.45	1.57	1.84	3.16
GAMMA,B=2.00	EXP	1.12	1.21	1.35	1.50	0.85	1.02	1.11	1.32
GAMMA,B=4.00	EXP	0.96	1.04	1.13	1.33	0.63	0.74	0.88	1.05
GAMMA,B=8.00	EXP	0.81	1.49	1.00	1.11	0.49	0.51	0.67	0.86
BATHTUB,P=0.15	EXP	1.27	1.40	1.48	2.08	1.21	1.29	1.41	1.90
BATHTUB,P=0.25	EXP	1.39	1.44	1.52	1.94	1.33	1.44	1.50	1.94
BATHTUB,P=0.40	EXP	1.49	1.54	1.61	1.89	1.42	1.60	1.59	1.92
BATHTUB,P=0.60	EXP	1.46	1.56	1.60	1.85	1.44	1.54	1.57	1.82
LN LOGISTIC,B= 2.	EXP	0.46	0.59	0.89	1.46	0.23	0.34	0.64	1.08
LN LOGISTIC,B= 3.	EXP	0.46	0.54	0.75	0.99	0.23	0.29	0.46	0.69
LN LOGISTIC,B= 4.	EXP	0.44	0.52	0.67	0.87	0.21	0.26	0.37	0.58

TABLE 6 - LOG LOGISTIC MLE

DISTRIBUTION OF DEATHS	CENSORING DISTRIBUTION	SAMPLE SIZE 25 PERCENT CENSORED 0%	10%	30%	50%	SAMPLE SIZE 50 PERCENT CENSORED 0%	10%	30%	50%
BATHTUB,P=0.60	EXP	1.14	1.16	1.23	1.34	0.87	0.87	0.87	0.95
EXPONENTIAL	EXP	1.32	1.37	1.49	1.85	1.12	1.16	1.16	1.42
WEIBULL,B=0.50	EXP	1.17	1.28	1.54	3.14	1.05	1.21	1.15	2.76
WEIBULL,B=2.00	EXP	1.29	1.32	1.43	1.55	1.14	1.17	1.24	1.37
WEIBULL,B=4.00	EXP	1.30	1.34	1.37	1.46	1.14	1.18	1.25	1.34
WEIBULL,B=8.00	EXP	1.34	1.40	1.45	1.51	1.16	1.20	1.25	1.37
WEIBULL,B=0.50	WEI		1.32	1.27	1.99		0.99	1.02	1.35
WEIBULL,B=2.00	WEI		1.31	1.41	1.69		1.16	1.21	1.43
WEIBULL,B=4.00	WEI		1.36	1.45	1.59		1.18	1.26	1.43
WEIBULL,B=8.00	WEI		1.42	1.47	1.62		1.17	1.33	1.39
NORMAL,σ=0.05	EXP	1.97	1.94	1.96	1.47	1.97	1.92	1.88	2.01
NORMAL,σ=0.10	EXP	1.72	1.75	1.84	1.39	1.65	1.72	1.69	1.72
NORMAL,σ=0.15	EXP	1.60	1.62	1.64	1.58	1.49	1.51	1.56	1.66
LN NORMAL,S=.37	EXP	1.47	1.49	1.53	1.54	1.43	1.46	1.49	1.64
LN NORMAL,S=.51	EXP	1.44	1.46	1.50	1.70	1.44	1.45	1.49	1.67
LN NORMAL,S=.61	EXP	1.44	1.47	1.58	1.79	1.44	1.50	1.58	1.79
EP,B=0.25	EXP	1.00	0.83	1.14	2.41	0.65	0.51	0.75	2.24
EP,B=0.50	EXP	1.16	1.11	1.08	1.50	0.83	0.75	0.72	1.10
EP,B=1.00	EXP	1.10	1.15	1.24	1.45	0.87	0.87	0.94	0.94
EP,B=8.00	EXP	1.21	1.17	1.33	1.20	0.85	0.92	1.06	1.16
EP,B=0.25	EP		0.94	0.90	1.15		0.59	0.50	0.67
EP,B=0.50	EP		1.12	1.10	1.31		0.81	0.72	0.87
EP,B=1.00	EP		1.11	1.21	1.31		0.86	0.86	0.99
EP,B=8.00	EP		1.18	1.31	1.03		0.93	1.04	0.97
PARETO,B=1.00	EXP	1.56	1.50	1.82	3.12	1.50	1.49	1.70	2.83
PARETO,B=2.00	EXP	1.47	1.50	1.71	2.48	1.37	1.35	1.41	2.05
PARETO,B=4.00	EXP	1.42	1.48	1.58	2.15	1.29	1.27	1.29	1.72
GOMPERTZ,B=0.50	EXP	1.14	1.18	1.23	1.41	0.90	0.93	0.93	1.02
GOMPERTZ,B=1.00	EXP	1.04	1.04	1.14	1.24	0.76	0.76	0.81	0.88
GOMPERTZ,B=2.00	EXP	0.91	0.95	1.01	1.09	0.67	0.69	0.72	0.77
GAMMA,B=0.50	EXP	1.21	1.11	1.12	1.82	0.87	0.78	0.73	1.19
GAMMA,B=2.00	EXP	1.44	1.41	1.57	1.78	1.30	1.29	1.44	1.60
GAMMA,B=4.00	EXP	1.44	1.49	1.57	1.65	1.39	1.36	1.51	1.62
GAMMA,B=8.00	EXP	1.46	1.04	1.55	1.58	1.43	1.46	1.51	1.62
BATHTUB,P=0.15	EXP	1.37	1.39	1.47	2.00	1.18	1.20	1.22	1.54
BATHTUB,P=0.25	EXP	1.32	1.39	1.51	1.89	1.12	1.17	1.22	1.54
BATHTUB,P=0.40	EXP	1.21	1.29	1.39	1.71	0.99	1.04	1.08	1.27
LN LOGISTIC,B= 2.	EXP	1.46	1.52	1.81	2.22	1.48	1.57	1.65	2.36
LN LOGISTIC,B= 3.	EXP	1.47	1.52	1.56	1.81	1.48	1.48	1.56	1.75
LN LOGISTIC,B= 4.	EXP	1.50	1.50	1.55	1.59	1.50	1.51	1.55	1.65

TABLE 7 - PARETO MLE

DISTRIBUTION OF DEATHS	CENSORING DISTRIBUTION	SAMPLE SIZE 25 PERCENT CENSORED 0%	10%	30%	50%	SAMPLE SIZE 50 PERCENT CENSORED 0%	10%	30%	50%
EXPONENTIAL	EXP	0.38	0.39	0.48	0.82	0.20	0.22	0.27	0.53
WEIBULL,B=0.50	EXP	1.23	1.21	1.07	1.81	1.00	1.23	0.91	1.82
WEIBULL,B=2.00	EXP	0.17	0.18	0.23	0.34	0.09	0.10	0.11	0.17
WEIBULL,B=4.00	EXP	0.02	0.03	0.03	0.05	0.01	0.01	0.02	0.02
WEIBULL,B=8.00	EXP	0.01	0.01	0.01	0.01	0.00	0.00	0.01	0.01
WEIBULL,B=0.50	WEI		1.32	1.09	1.17		1.09	0.93	0.86
WEIBULL,B=2.00	WEI		0.18	0.23	0.39		0.10	0.12	0.21
WEIBULL,B=4.00	WEI		0.03	0.03	0.05		0.01	0.02	0.03
WEIBULL,B=8.00	WEI		0.01	0.01	0.01		0.00	0.00	0.01
NORMAL,σ=0.05	EXP	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
NORMAL,σ=0.10	EXP	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00
NORMAL,σ=0.15	EXP	0.01	0.01	0.01	0.01	0.00	0.00	0.01	0.01
LN NORMAL,S=.37	EXP	0.03	0.03	0.04	0.06	0.02	0.02	0.02	0.03
LN NORMAL,S=.51	EXP	0.07	0.08	0.10	0.13	0.04	0.04	0.05	0.07
LN NORMAL,S=.61	EXP	0.14	0.15	0.18	0.25	0.08	0.08	0.10	0.13
EP,B=0.25	EXP	1.01	0.74	0.86	3.35	0.62	0.43	0.59	3.96
EP,B=-0.50	EXP	0.41	0.52	0.66	1.22	0.46	0.57	0.72	1.28
EP,B=1.00	EXP	0.36	0.39	0.52	0.79	0.22	0.24	0.31	0.49
EP,B=8.00	EXP	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.01
EP,B=0.25	EP		0.87	0.71	0.80		0.52	0.40	0.51
EP,B=0.50	EP		0.47	0.63	0.88		0.66	0.72	0.93
EP,B=1.00	EP		0.39	0.51	0.86		0.24	0.32	0.55
EP,B=8.00	EP		0.01	0.01	0.01		0.00	0.00	0.00
PARETO,B=1.00	EXP	0.30	0.31	0.29	0.54	0.77	0.47	0.30	0.40
PARETO,B=2.00	EXP	0.32	0.31	0.36	0.55	0.44	0.39	0.26	0.43
PARETO,B=4.00	EXP	0.33	0.33	0.38	0.70	0.29	0.29	0.29	0.41
GOMPERTZ,B=0.50	EXP	0.39	0.41	0.55	0.87	0.23	0.26	0.31	0.54
GOMPERTZ,B=1.00	EXP	0.22	0.25	0.32	0.49	0.13	0.14	0.18	0.27
GOMPERTZ,B=2.00	EXP	0.07	0.08	0.09	0.14	0.04	0.04	0.05	0.07
GAMMA,B=0.50	EXP	0.51	0.63	0.69	1.22	0.86	0.75	0.88	1.38
GAMMA,B=2.00	EXP	0.28	0.31	0.38	0.53	0.15	0.17	0.21	0.31
GAMMA,B=4.00	EXP	0.07	0.08	0.09	0.13	0.04	0.04	0.05	0.07
GAMMA,B=8.00	EXP	0.03	0.25	0.04	0.05	0.01	0.01	0.02	0.02
BATHTUB,P=0.15	EXP	0.37	0.38	0.50	0.84	0.25	0.23	0.31	0.52
BATHTUB,P=0.25	EXP	0.36	0.37	0.50	0.80	0.19	0.23	0.32	0.53
BATHTUB,P=0.40	EXP	0.33	0.40	0.54	0.87	0.18	0.22	0.31	0.56
BATHTUB,P=0.60	EXP	0.36	0.40	0.56	0.86	0.21	0.22	0.30	0.53
LN LOGISTIC,B= 2.	EXP	0.26	0.25	0.29	0.49	0.16	0.15	0.16	0.25
LN LOGISTIC,B= 3.	EXP	0.12	0.13	0.16	0.21	0.07	0.07	0.08	0.12
LN LOGISTIC,B= 4.	EXP	0.05	0.05	0.06	0.08	0.03	0.03	0.03	0.04

TABLE 8 - GOMPERTZ MLE

DISTRIBUTION OF DEATHS	CENSORING DISTRIBUTION	SAMPLE SIZE 25 PERCENT CENSORED 0%	10%	30%	50%	SAMPLE SIZE 50 PERCENT CENSORED 0%	10%	30%	50%
EXPONENTIAL	EXP	1.41	1.44	1.57	2.16	1.57	1.52	1.63	2.24
WEIBULL,B=0.50	EXP	0.26	0.44	1.09	3.15	0.15	0.41	0.62	2.03
WEIBULL,B=2.00	EXP	1.20	1.22	1.27	1.37	0.95	0.99	1.06	1.22
WEIBULL,B=4.00	EXP	1.20	1.24	1.29	1.40	0.93	0.97	1.10	1.17
WEIBULL,B=8.00	EXP	1.50	1.52	1.61	1.63	1.35	1.40	1.47	1.54
WEIBULL,B=0.50	WEI		0.35	0.66	1.43		0.21	0.40	0.85
WEIBULL,B=2.00	WEI		1.23	1.34	1.61		0.99	1.14	1.50
WEIBULL,B=4.00	WEI		1.27	1.39	1.69		1.01	1.17	1.51
WEIBULL,B=8.00	WEI		1.54	1.56	1.79		1.36	1.51	1.69
NORMAL,σ=0.05	EXP	0.65	0.77	0.97	1.16	0.36	0.41	0.51	0.72
NORMAL,σ=0.10	EXP	0.78	0.82	1.05	1.35	0.46	0.51	0.66	0.89
NORMAL,σ=0.15	EXP	0.82	0.93	1.07	1.31	0.54	0.59	0.74	0.94
LN NORMAL,S=.37	EXP	0.47	0.52	0.66	0.86	0.24	0.29	0.38	0.53
LN NORMAL,S=.51	EXP	0.54	0.64	0.75	0.92	0.29	0.35	0.46	0.65
LN NORMAL,S=.61	EXP	0.66	0.70	0.83	1.04	0.37	0.42	0.55	0.76
EP,B=0.25	EXP	0.19	0.31	1.04	3.32	0.11	0.17	0.62	2.50
EP,B=0.50	EXP	0.87	0.94	1.23	2.43	0.62	0.66	0.88	1.92
EEP,B=1.00	EXP	1.48	1.45	1.53	1.71	1.49	1.47	1.58	1.67
EP,B=8.00	EXP	1.85	1.73	1.89	1.77	1.72	1.79	1.81	1.78
EP,B=0.25	EP		0.24	0.50	1.26		0.14	0.28	0.67
EP,B=0.50	EP		0.91	1.17	1.73		0.68	0.84	1.29
EP,B=1.00	EP		1.47	1.53	1.77		1.53	1.52	1.79
EP,B=8.00	EP		1.68	1.77	1.83		1.75	1.75	1.77
PARETO,B=1.00	EXP	0.33	0.68	1.42	3.16	0.22	0.51	1.07	2.45
PARETO,B=2.00	EXP	0.69	0.94	1.48	2.64	0.58	0.79	1.23	2.31
PARETO,B=4.00	EXP	1.00	1.22	1.55	2.46	0.94	1.12	1.41	2.24
GOMPERTZ,B=0.50	EXP	1.43	1.44	1.50	1.79	1.49	1.49	1.54	1.64
GOMPERTZ,B=1.00	EXP	1.47	1.48	1.51	1.57	1.46	1.50	1.55	1.62
GOMPERTZ,B=2.00	EXP	1.50	1.48	1.54	1.62	1.53	1.53	1.58	1.64
GAMMA,B=0.50	EXP	0.65	0.79	1.19	2.64	0.44	0.52	0.81	1.92
GAMMA,B=2.00	EXP	1.19	1.20	1.29	1.39	0.94	0.99	1.06	1.22
GAMMA,B=4.00	EXP	0.81	0.87	0.98	1.16	0.48	0.56	0.70	0.87
GAMMA,B=8.00	EXP	0.64	1.48	0.85	0.99	0.36	0.40	0.54	0.72
BATHTUB,P=0.15	EXP	1.36	1.43	1.52	2.27	1.52	1.45	1.56	2.18
BATHTUB,P=0.25	EXP	1.41	1.44	1.54	2.03	1.44	1.49	1.61	2.19
BATHTUB,P=0.40	EXP	1.45	1.46	1.56	1.98	1.49	1.52	1.58	2.02
BATHTUB,P=0.60	EXP	1.45	1.48	1.54	1.79	1.50	1.49	1.52	1.74
LN LOGISTIC,B= 2.	EXP	0.61	0.78	1.08	1.49	0.38	0.61	0.85	1.19
LN LOGISTIC,B= 3.	EXP	0.51	0.56	0.73	0.93	0.26	0.31	0.44	0.63
LN LOGISTIC,B= 4.	EXP	0.39	0.46	0.59	0.80	0.18	0.22	0.32	0.51

FIGURE 1 - **MSE OF MAXIMUM LIKELIHOOD ESTIMATORS OF SURVIVAL FOR THE EXPONENTIAL DISTRIBUTION WITH 30% CENSORING**

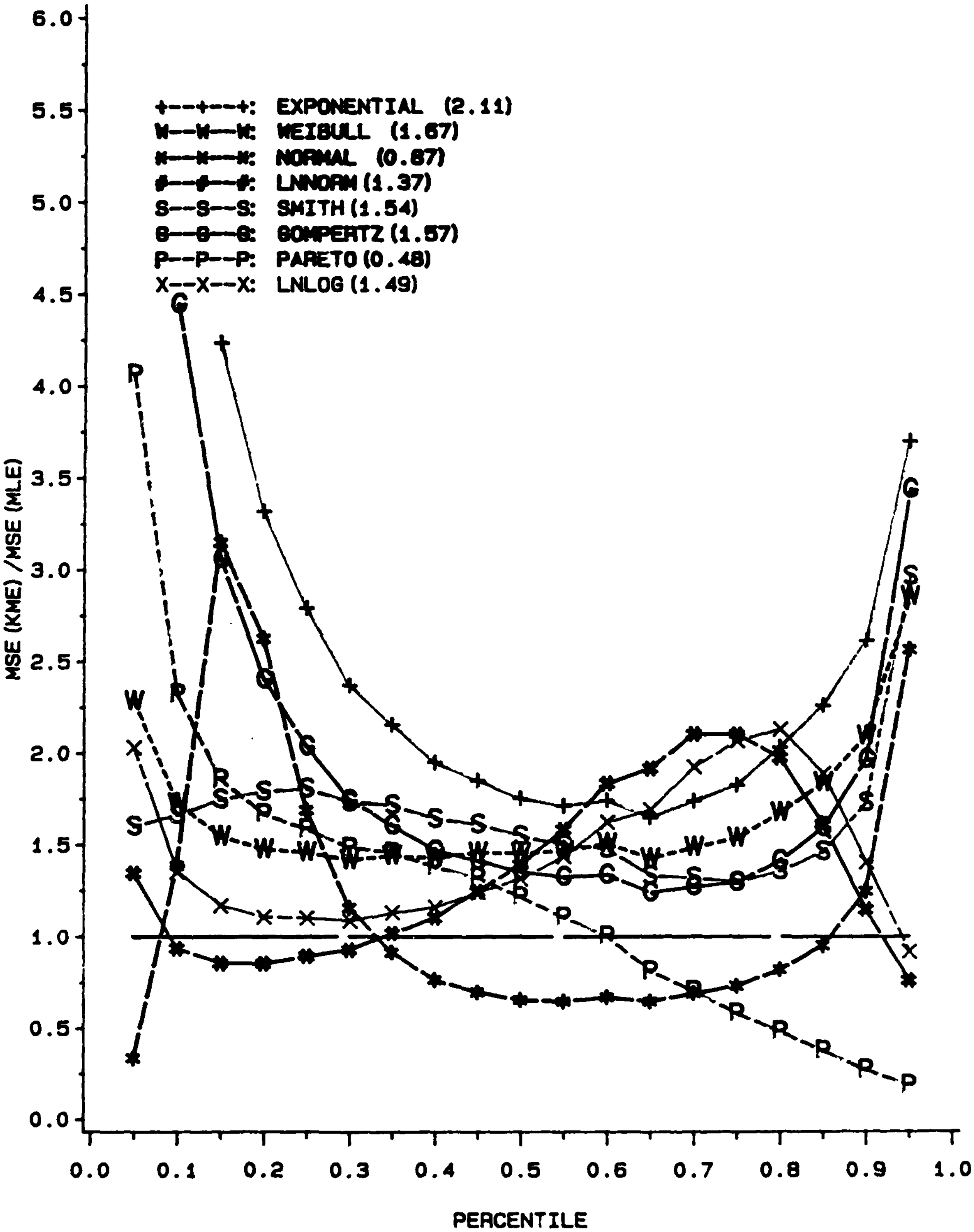


FIGURE 2 - **MSE OF MAXIMUM LIKELIHOOD ESTIMATORS OF SURVIVAL FOR THE GAMMA DISTRIBUTION WITH B=0.5, 30% CENSORING**

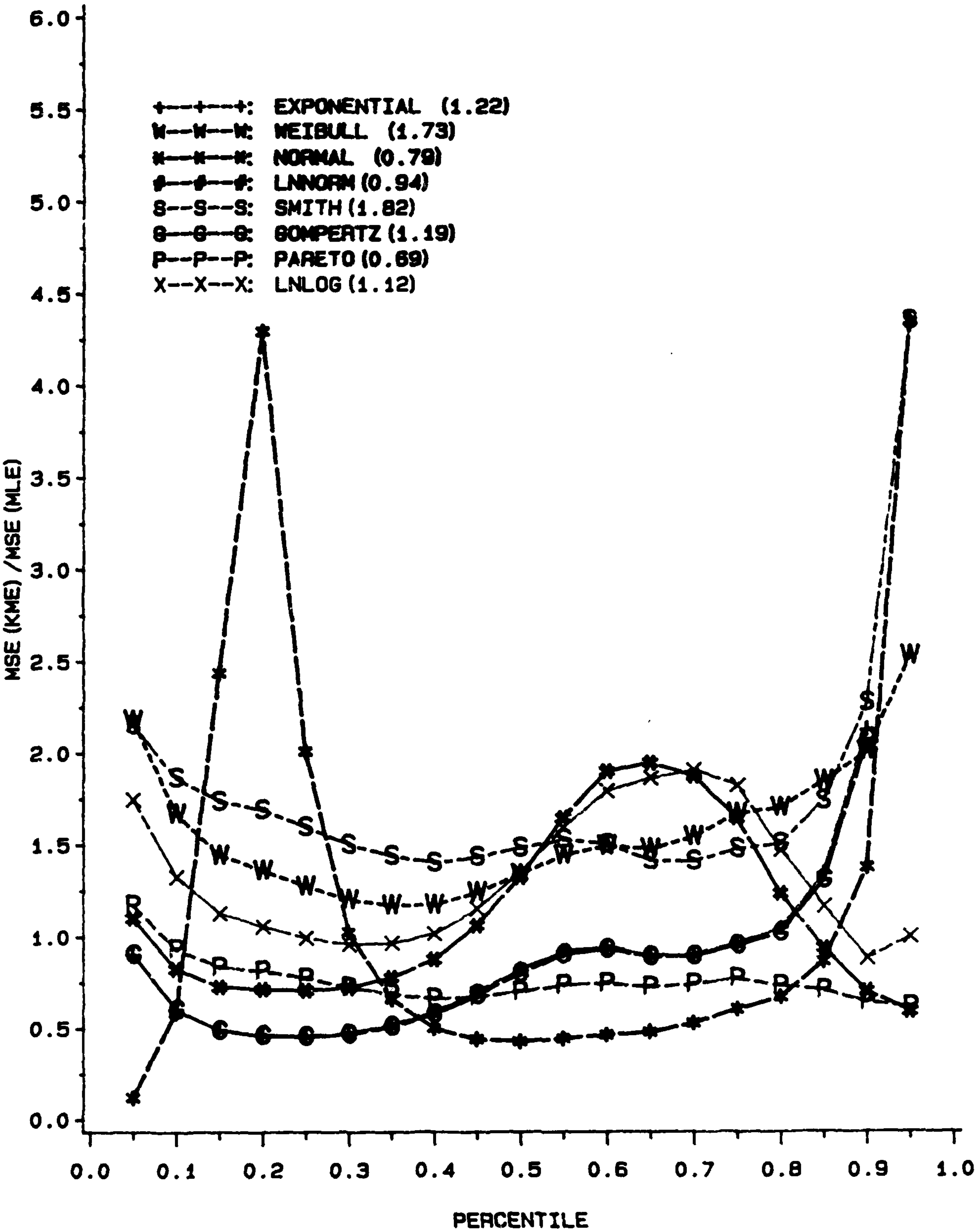


FIGURE 3 - **MSE OF MAXIMUM LIKELIHOOD ESTIMATORS OF SURVIVAL FOR THE WEIBULL DISTRIBUTION WITH B=8.0, 30% CENSORING**

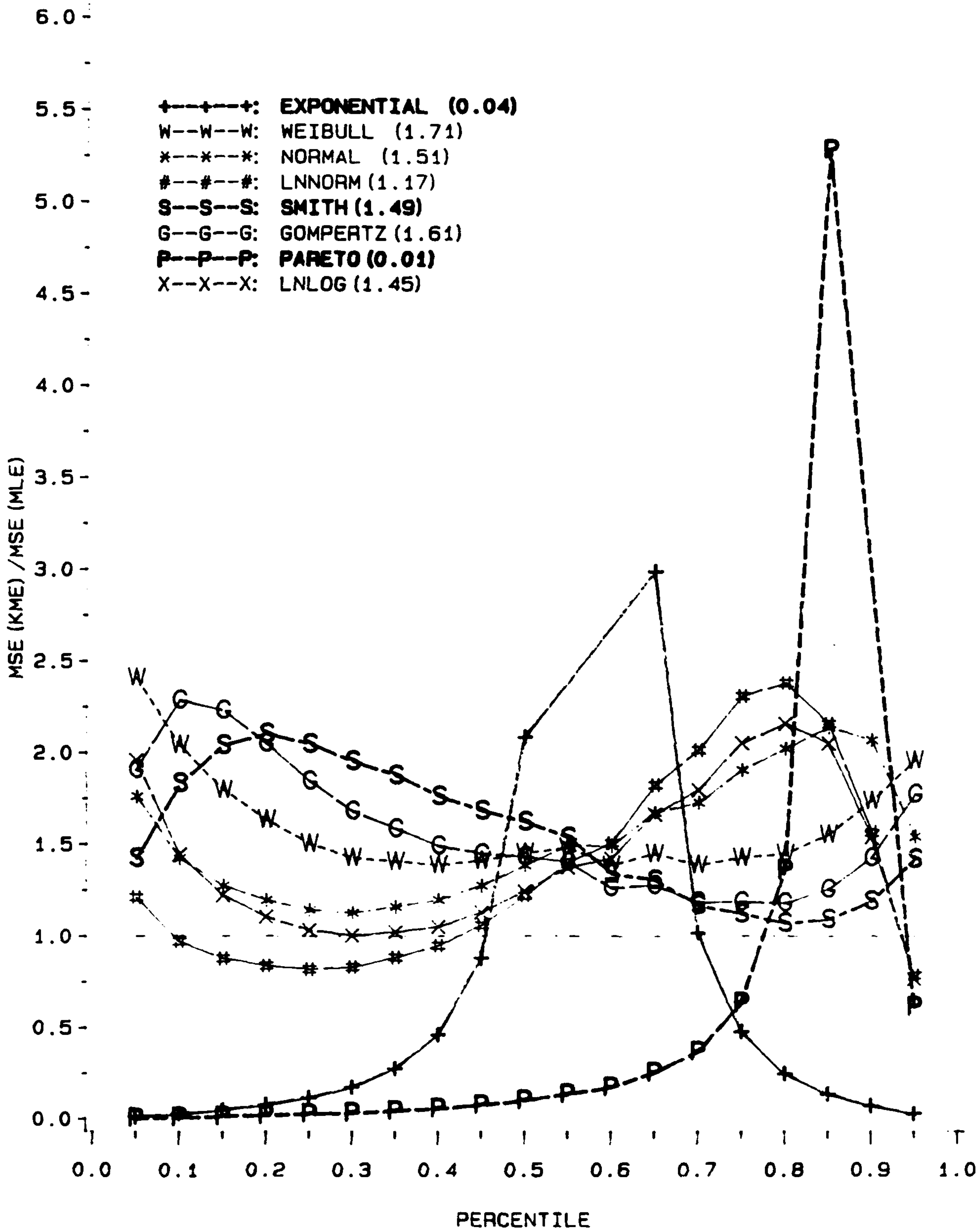


FIGURE 4 - **MSE OF MAXIMUM LIKELIHOOD ESTIMATORS OF SURVIVAL FOR THE GOMPERTZ DISTRIBUTION WITH B=2, 30% CENSORING**

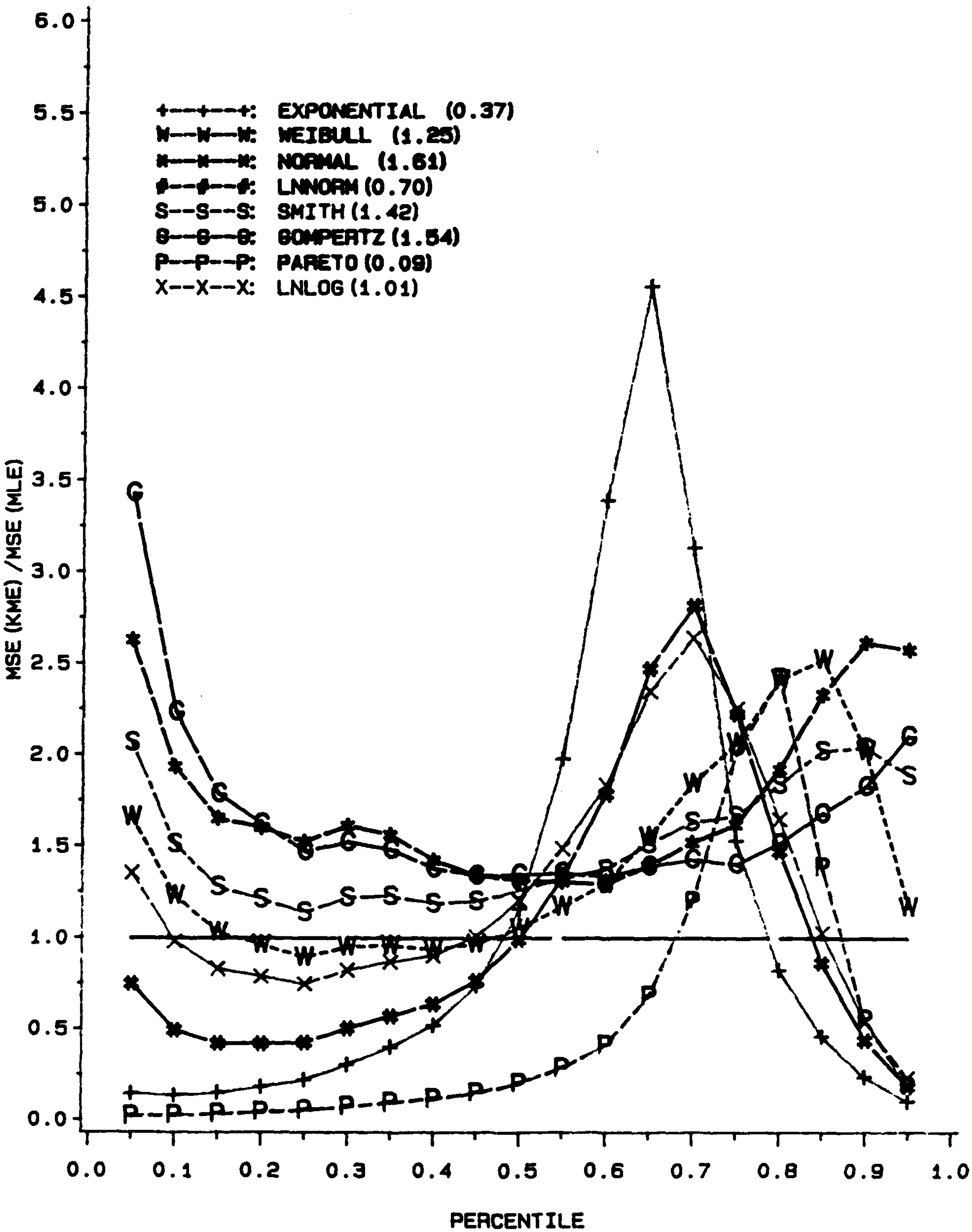


FIGURE 5 - **MSE OF MAXIMUM LIKELIHOOD ESTIMATORS OF SURVIVAL FOR THE LOG-LOGISTIC DISTRIBUTION WITH B=3, 30% CENSORING**

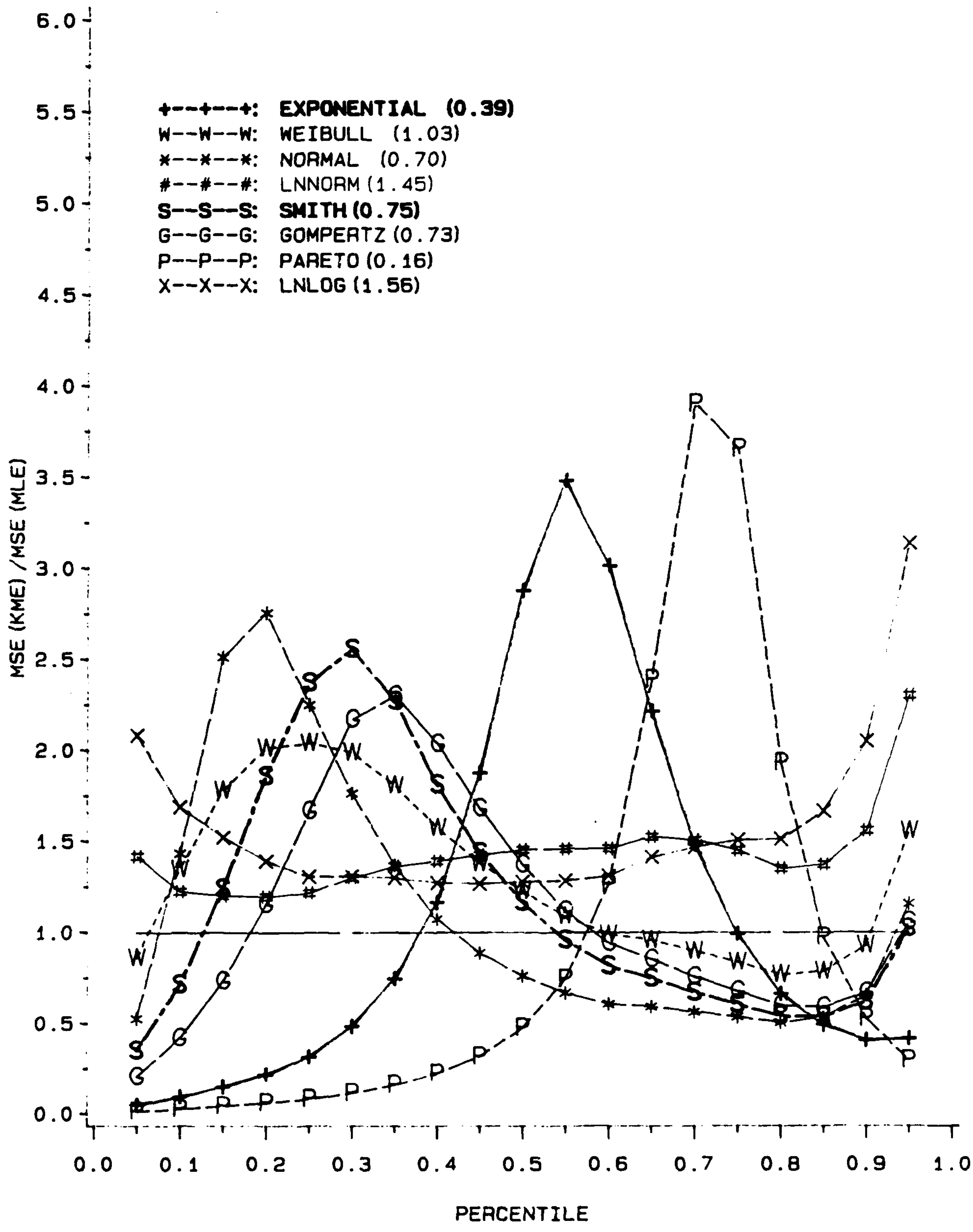


FIGURE 6 - **MSE OF MAXIMUM LIKELIHOOD ESTIMATORS OF SURVIVAL FOR THE BATHTUB DISTRIBUTION WITH P=0.6, 30% CENSORING**

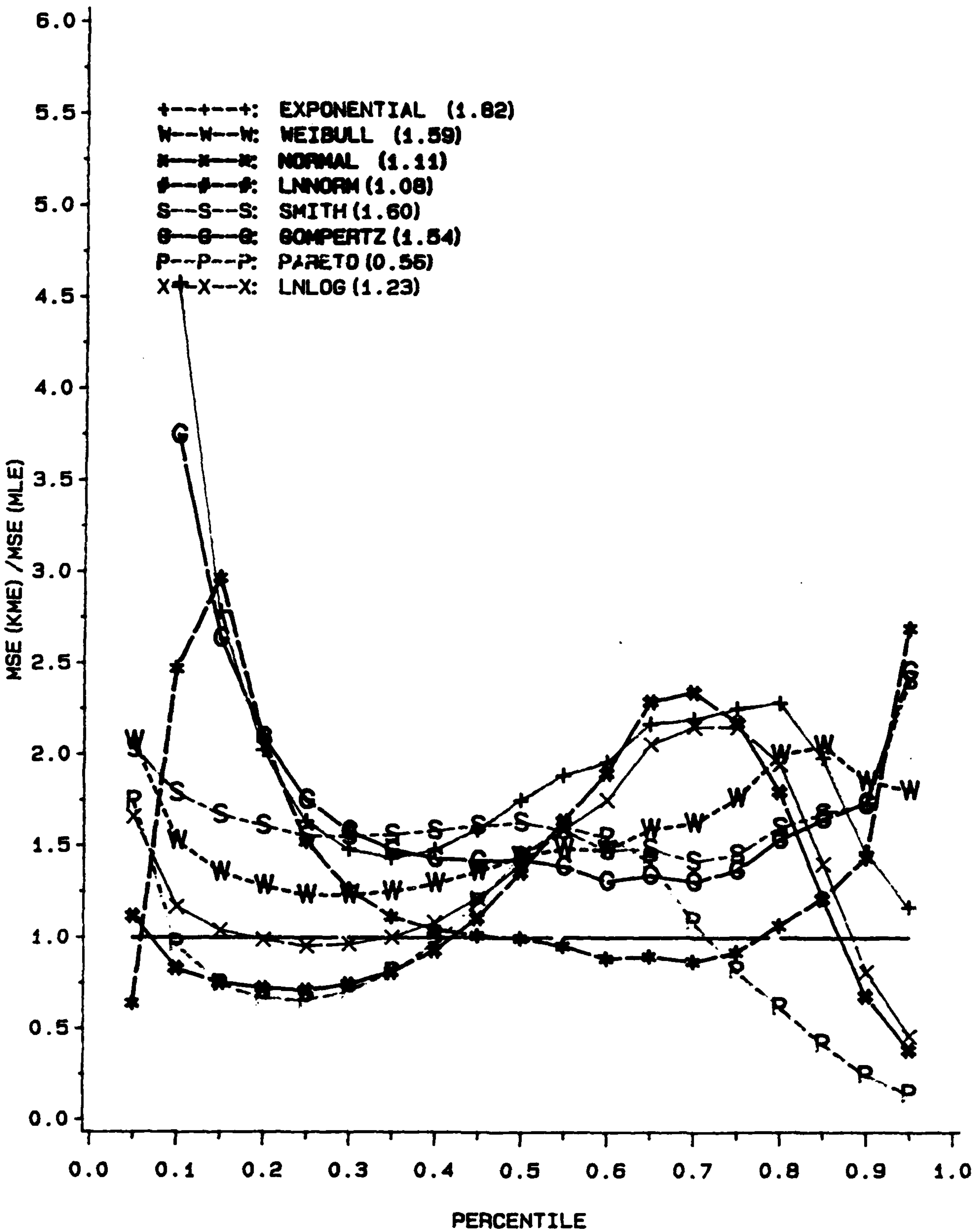


FIGURE 7 - **MSE OF MAXIMUM LIKELIHOOD ESTIMATORS OF SURVIVAL FOR THE SMITH DISTRIBUTION WITH B=0.25, 30% CENSORING**